Equity Premium Forecasts with an Unknown Number of Structural Breaks

GEORGE BULKLEY, DAVID S. LESLIE AND SIMON C. SMITH*

ABSTRACT

Estimation of models with structural breaks usually assumes a pre-specified number of breaks. Previous models which do allow an endogenously determined number of breaks require a simple structural model, and rarely allow for information transfer across the break. We introduce a methodology that allows the number of breaks to be determined endogenously and including an economically-motivated model of transition regimes between each break. We demonstrate the usefulness of our approach for forecasts of the equity premium. We find the demonstrated success of the historical average can be improved upon by an economic model with theory informed priors estimated using our methodology.

JEL classification: G10, C11, C15, C53.

---

*Bulkley is based in the Department of Accounting and Finance at the University of Bristol. Address: 15-19 Tyndalls Park Road, Clifton, BS8 1TU. Email: george.bulkley@bristol.ac.uk. Telephone: +44117 39 41478. Leslie is based in the Department of Mathematics and Statistics at Lancaster University. Address: Bailrigg Campus, Lancaster, LA1 4YF. Smith (corresponding author) is based in USC Dornsife INET, Department of Economics, University of Southern California. Address: Institute for New Economic Thinking, Department of Economics, USC, 3620 South Vermont Ave., CA, 90089-0253, USA. Much of the work was completed while Smith was at University of Bristol and Lancaster University. Special thanks must go to the editor (Fabio Trojani) and two anonymous referees. We would also like to thank Paolo Giordani, John Maheu, Mark Shackleton, and Jon Temple for helpful suggestions, discussions, and comments, and seminar participants at Royal Economic Society (Doctoral Symposium), Bristol Econometric Study Group 2015, 2nd Young Finance Scholars' Conference, World Finance Conference (Buenos Aires), and the 10th Annual RCEA Bayesian Econometric Workshop. All errors are our own. Smith is grateful for financial support from the ESRC.

1. Introduction

There is mounting evidence that many macroeconomic and financial time series are subject to occasional shifts in parameters known as structural breaks (Stock and Watson, 1996; Ang and Bekaert, 2002).[1] Ignoring such breaks can be very costly for forecasting. Clements and Hendry (1998) suggest that ignoring breaks is the primary cause of low forecasting power. Indeed several recent works demonstrate that incorporating structural breaks can significantly improve the forecasting power of econometric models (Pesaran and Timmermann, 2002; Pesaran, Pettenuzzo, and Timmermann, 2006; Koop and Potter, 2007; Giacomini and Rossi, 2009; Inoue and Rossi, 2011).

In this paper we introduce a Bayesian multiple changepoint model where the number of breaks is determined endogenously using a flexible and generic approach. We develop a reversible jump Markov chain Monte Carlo algorithm (Green, 1995) that is applicable to highly structured Bayesian multiple changepoint models that cannot be easily estimated using existing approaches. We demonstrate how our methodology can be employed in the context of one such model that estimates the path of the equity premium over a long sample by formally modelling the multiple breakpoint process (Pastor and Stambaugh, 2001). An important feature of this model is the assumption that successive shifts in the state variable give rise to transitions regimes that reflect learning about these breaks. We might expect transition regimes in other models where a series of interest reflects learning, for example long yields that reflect learning about shifts in short yields, or exchange rates that reflect learning about shifts in fundamentals. The structure of these breaks and transition regimes are informed by economic theory through the specification of the prior distributions. Fur-

[1] A small handful off frequentist changepoint approaches include Andrews (1993), Bai (1997), Bai and Perron (1998), Bai and Perron (2003), and Chen and Hong (2012).

thermore, the Bayesian methodology also allows the uncertainty surrounding the parameters of the model, including the timing of breaks, to be incorporated into the equity premium estimates. In their estimation Pastor and Stambaugh (2001) use the algorithm of Chib (1998), which is restricted by the assumption that the number of breaks needs to be pre-specified by the user. The consequence is that the full advantages of this appealing economic model cannot be exploited. The merit of our reversible jump Markov chain Monte Carlo approach is that it can be used to estimate these highly structured models with an endogenously determined number of change points.

Previous Bayesian multiple changepoint models include that introduced by Pesaran et al. (2006) who extend the framework of Chib (1998) to allow for out-of-sample breaks, but still need to pre-specify the number of in-sample breaks.[2] Koop and Potter (2007) made a significant advance by extending the algorithm of Chib (1998) so that the number of in-sample breaks may vary.[3]

The sampling approaches developed by Chib (1998) and Koop and Potter (2007) update either the break locations or the parameters at any particular time. These approaches can (i) have difficulty mixing and (ii) be sensitive to how the chain is initialised (Fearnhead, 2006). In contrast we propose a fitting scheme which updates both break locations and parameters simultaneously, increasing the probability that the estimator converges to the true data generating process as it does not have to pass through potentially implausible intermediate moves, and is less sensitive to chain initialisation. Furthermore, our fitting method is not specific to any particular prior on the regime durations, unlike the methods of Chib (1998) and Koop and Potter (2007). Any prior can be easily specified by making small

[2]Other Bayesian changepint studies include Chernoff and Zacks (1964), Carlin, Gelfand, and Smith (1992), Stephens (1994), and Ko, Chong, Ghosh, et al. (2015) to name but a few.

[3]See Bauwens, Koop, Korobilis, and Rombouts (2015) for an application of the method of Koop and Potter (2007) to macroeconomic forecasting.

adjustments to the fitting algorithm, and these adjustments will not have significant impact on the running time of the method; this flexibility is in stark contrast with the prior-specific approach of Koop and Potter (2007) which incurs a significant computational penalty in switching from geometric to Poisson regime durations.

Giordani and Kohn (2008) introduce an efficient methodology for estimating conditionally Gaussian state-space models. Unfortunately an economically-informed structural breaks model such as Pastor and Stambaugh (2001) cannot be cast in their framework, since the dependence structure through a transition regime that represents a gradual structural break precludes formulation as a Markovian state space model, even when conditioning on latent variables. For similar reasons, the methodology of Maheu and McCurdy (2009) cannot be applied in a more complex model like that of Pastor and Stambaugh (2001); their methodology also relies on the fact that no information is transmitted across a break. More recently, the infinite Hidden Markov Model approach of Fox, Sudderth, Jordan, and Willsky (2011) (see also Song (2014)) has been proposed that approximately nests the multiple change-point approach and the regime-switching framework of Hamilton (1989). However again this approach cannot accommodate the idea of transition regimes.

Giordani, Kohn, and van Dijk (2007) and Koop and Potter (2010) (see also Lundbergh, Teräsvirta, and Van Dijk (2003), and Bec, Rahbek, and Shephard (2008)) develop frameworks for estimating smooth transition autoregressive models. Since we are interested in applying the model of Pastor and Stambaugh (2001) in a recursive forecasting scheme we develop a framework for estimating a model with gradual shifts in mean (and variance) as opposed to a smooth transition autoregressive model.

In order to evaluate the benefits of our methodology we first revisit the full sample equity premium estimates of Pastor and Stambaugh (2001). We find that fixing the number of breaks to 15 relative to incorporating the uncertainty surrounding the number of breaks into the estimates of the equity premium tends to underestimate the premium. It also results in estimation errors as high as 2.5 percentage points around the time of the Wall Street Crash which is characterised as the most unstable period in our sample. Unstable periods such as the oil price shocks of the 1970s and the Global Financial Crisis also exhibit large estimation errors. The root-squared difference in estimates averages 0.78% across the sample.

Second, we apply our methodology to the problem of forecasting the U.S. equity premium. Since this time series has been found to be subject to structural breaks (Kim, Morley, and Nelson, 2005; Rapach, Strauss, and Zhou, 2009) a methodology that is better able to identify these breaks should yield improved forecasts compared to conventional ad hoc techniques that allow for breaks. The historical average has delivered very strong forecasts of the equity premium outperforming a range of economic models (Goyal and Welch, 2008). This success of the historic average is particularly surprising given the evidence of breaks in the equity premium. If our method has real merits we should expect it to deliver superior forecasts in this application. In order to judge whether our methodology yields tangible benefits we test whether a mean variance investor could have made economic profits by employing our methodology.

We find that forecasts obtained from a theoretically motivated Bayesian univariate model in which stochastic structural breaks are explicitly modelled consistently beat the historical average in the full sample 1927 to 2013 and in all of the three subsamples that we examine. Moreover a risk averse mean-variance investor who employed a simple trading rule that varies the fraction of their portfolio invested in equities with the forecast premium could have made economically significant profits in the full sample and each of the three subsamples. We find that a range of ad hoc methods for dealing with breaks such as rolling windows and an exponential smoothing model rarely outperform the historical average.

The remainder of this paper is structured as follows. Section 2 introduces our new model, a modification of Pastor and Stambaugh (2001)'s model. Section 3 describes an empirical test that compares the predictive abilities of several models. Section 4 presents and discusses the empirical results. Section 5 concludes.

2. Methodology

In this section we introduce the main structural breaks model that we estimate. Our base model is that of Pastor and Stambaugh (2001). This Bayesian model incorporates transition regimes and allows the model to be informed by economic theory through the specification of prior distributions.

Within our framework a structural break occurs gradually not instantly. A regime is a period separated by two successive changepoints. Since the transition from one stable regime to the next occurs gradually, each structural break is characterised by two changepoints. The gradual break that occurs between these two successive changepoints is called a transition regime (TR), during which the equity premium is transiting from one level to another. Each TR separates two neighbouring stable regimes (SRs), during which the distribution of returns is stable.

Allowing for gradual breaks means a negative prior association can be specified so that high returns are expected during a TR in which the equity premium is transiting to a lower value and vice versa. The strength of this negative relation is specified through the prior. The ratio of the equity premium to variance, the 'price of risk', in the current regime helps to estimate the current equity premium given the current regime's estimated variance. The dispersion in the prices of risk across regimes reflects the prior belief regarding the strength of the relation between volatility and the equity premium. It is also assumed that the equity premium is unlikely to be subject to large shifts. A prior is specified that corresponds to this assumption, placing most of the prior weight on relatively small shifts in the equity premium. The equity premium is truncated at zero because risk averse investors must be compensated for taking on increased risk and therefore nonpositive values of the equity premium should be ruled out on theoretical grounds.

A Bayesian framework is also natural for a recursive forecasting exercise in the presence of multiple structural breaks because it enables the uncertainty surrounding both the number and timing of breaks to be incorporated into the parameter estimates. A frequentist approach to modelling structural breaks consists of two stages. First, the break locations are estimated, and then the parameters of the model are estimated conditional on the estimated break locations. To show this approach more formally let θ denote the parameter vector, x denote the data, and q denote the changepoint locations. A frequentist estimates the changepoint locations, $\hat{q}$, and then estimates the parameter vector, θ, conditional on the data, x, and the estimated changepoint locations, $\hat{q}$. Assuming that the estimated changepoint locations are the true locations when estimating the parameters could however lead to estimation risk as it ignores any errors in the estimates, $\hat{q}$. This assumption could compromise parameter estimates in finite samples.

The Bayesian approach of Pastor and Stambaugh (2001) enables them to draw the posterior distribution of the changepoint locations conditional on the data, $p(q \mid x)$, and marginalise q when estimating θ

$$p(\theta \mid x) = \int_q p(\theta \mid x, q) p(q \mid x) dq, \tag{1}$$

thereby incorporating the uncertainty surrounding the changepoint locations into their estimates of θ. However, Pastor and Stambaugh (2001) are constrained by the algorithm of Chib (1998) to fix the number of breaks in advance. Let K be the assumed number of breaks in the sample. Pastor and Stambaugh (2001) perform their inference conditional on a fixed number of breaks and therefore K has been suppressed in the above notation. Explicitly they estimate θ as

$$p(\theta \mid x, K) = \int_q p(\theta \mid x, q, K) p(q \mid x, K) dq, \tag{2}$$

in which they set $K = 15$.

The key technical innovation of this paper compared with Pastor and Stambaugh (2001) is the development of a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm that allows us to approximate the posterior distribution of the number of breaks (Green, 1995). This has two main benefits. First, it enables the number of breaks to be inferred automatically from the data for each estimation period in the recursive forecasting exercise. Determining the number of breaks automatically enables the model of Pastor and Stambaugh (2001), and potentially many other complex Bayesian models, to be easily applied in a recursive forecasting exercise. Second, by approximating $p(K \mid x)$ we can go one step further than Pastor and Stambaugh (2001) by also marginalising K when estimating θ

$$p(\theta \mid x) = \int_K \int_q p(\theta \mid x, q, K) p(q \mid x, K) p(K \mid x) dq dK, \tag{3}$$

thereby incorporating not only the uncertainty surrounding the timing of the breaks but also the uncertainty surrounding the number of breaks in our estimates of θ.

The model is estimated in exactly the same way as in Pastor and Stambaugh (2001) but with the addition of a RJMCMC step that enables the addition or removal of a changepoint, and the removal of Chib (1998)'s changepoint methodology that is no longer necessary. Full details are given in the Appendix. We now briefly revisit the model of Pastor and Stambaugh (2001), indicating where we make modifications.

2.1. Pastor and Stambaugh's model

Let $x = (x_1, \ldots, x_T)$ denote the vector of T observations of excess market returns in the data, in which x_t denotes the excess market return at time t. Conditional on the number of structural breaks, K, which is fixed, the sample period is split into $2K + 1$ regimes. There are $K + 1$ stable regimes (SRs) that are separated by the remaining K transition regimes (TRs). The regimes alternate in order beginning and ending with a SR. The probability distribution of excess market returns is stable in SRs but is undergoing change during each TR.

Let $q = (q_1, \ldots, q_{2K})$ denote a vector of changepoints, the times at which one regime transits to the next. For convenience, assume $q_0 = 0$. The first time point in the ith SR is $q_{2i-2} + 1$ and the final time point is q_{2i-1}. The first time point in the jth TR is $q_{2j-1} + 1$ and the final time point is q_{2j}. Denoting the duration of regime k as $l_k = q_k - q_{k-1}$ the duration of the ith SR and jth TR are l_{2i-1} and l_{2j}, respectively. During each SR_i the excess market returns are assumed to be normally distributed with a regime specific mean μ_i and variance σ_i^2

$$x_t \sim \mathcal{N}(\mu_i, \sigma_i^2), \qquad t \in SR_i, \qquad i = 1, \ldots, K + 1. \tag{4}$$

Let $\mu = (\mu_1, \ldots, \mu_{K+1})$ denote the vector of equity premiums and $\sigma_{SR}^2 = (\sigma_1^2, \ldots, \sigma_{K+1}^2)$ denote the volatilities in the SRs.

The mean excess return during each TR consists of two components. The first is the average of the mean excess returns in the two neighbouring SRs. This component represents the unconditional expected return during the TR. The second component represents the conditional expected return during the TR by conditioning on the shift in the mean excess return, $\mu_{j+1}-\mu_j$. The second component therefore allows a prior to be specified that reflects Pastor and Stambaugh (2001)'s belief that one expects to see high returns during TRs in which the premium falls and low returns during TRs in which the premium rises. The likelihood of x_t during each TR_j is

$$x_t \sim \mathcal{N}\left(\frac{\mu_j+\mu_{j+1}}{2}+b_j(\mu_{j+1}-\mu_j),\sigma^2_{j,j+1}\right), \qquad t\in TR_j, \qquad j=1,\dots,K. \tag{5}$$

Let $\sigma^2_{TR}=(\sigma^2_{1,2}\dots,\sigma^2_{K,K+1})$ denote the volatilities in the TRs and $b=(b_1,\dots,b_K)$ denote the negative discount conditions. The likelihood function is

$$\begin{aligned}
&p(x \mid \mu,\sigma^2_{SR},\sigma^2_{TR},b,q,K)\\
&\quad\propto\left(\prod_{i=1}^{K+1}\frac{1}{\sigma_i^{l_{2i-1}}}\right)\exp\left\{-\frac{1}{2}\sum_{i=1}^{K+1}\sum_{t=q_{2i-2}+1}^{q_{2i-1}}\frac{(x_t-\mu_i)^2}{\sigma_i^2}\right\}\\
&\quad\quad\times\left(\prod_{j=1}^{K}\frac{1}{\sigma_{j,j+1}^{l_{2j}}}\right)\exp\left\{-\frac{1}{2}\sum_{j=1}^{K}\sum_{t=q_{2j-1}+1}^{q_{2j}}\frac{\left(x_t-\left(\frac{\mu_j+\mu_{j+1}}{2}+b_j\left(\mu_{j+1}-\mu_j\right)\right)^2\right)}{\sigma^2_{j,j+1}}\right\}.
\end{aligned} \tag{6}$$

It is difficult to perform meaningful inference over the full joint likelihood in a frequentist setting therefore Pastor and Stambaugh (2001) suggest a Bayesian approach.

2.2. Prior beliefs

A Bayesian approach combines information in the data transmitted through the likelihood function with prior information about the parameters of the model provided by the researcher. This approach enables economic theory to inform the model through the prior distributions. Throughout, we use exactly the same prior specifications as Pastor and Stambaugh (2001). We therefore do not discuss the choices here, but for completeness note what the prior specifications are in the remainder of this section. The only point where we could differ is in the autoregressive structure placed on the equity premium values in the SRs.

2.2.1. Beliefs about regime durations.

The duration of the kth regime for $k = 1, \ldots, 2K$ follows a geometric distribution conditional on the nontransition probability of that regime, $p_{k,k}$, and the current estimated number of breaks, K

$$p(l_k \mid p_{k,k}) = p_{k,k}^{l_k - 1}(1 - p_{k,k}), \tag{7}$$

in which $p_{k,k}$ has a beta prior[4]

$$p(p_{k,k}) = \frac{\Gamma(a_k + c_k)}{\Gamma(a_k)\Gamma(c_k)} p_{k,k}^{a_k - 1}(1 - p_{k,k})^{c_k - 1}. \tag{8}$$

The marginal prior distribution of the duration of the kth regime, $p(l_k)$, for regimes $k = 1, \ldots, 2K$ can then be calculated as

$$\begin{aligned} p(l_k) &= \int_0^1 p(l_k \mid p_{k,k}) p(p_{k,k}) dp_{k,k}, \\ p(l_k) &= \frac{c_k \Gamma(a_k + c_k)\Gamma(l_k + a_k - 1)}{\Gamma(a_k)\Gamma(a_k + l_k + c_k)}, \qquad k \in 1, \ldots, 2K. \end{aligned}$$

Chib (1998)'s method forces Pastor and Stambaugh (2001) to constrain the nontransition probability of the final regime, $p_{2K+1,2K+1}$, to be equal to 1. We do not assume that the final

[4]For simplicity, we suppress the conditioning notation on a_k and c_k throughout.

regime ends at time T so q_{2K+1} may occur beyond the end of the sample. The probability of observing l_{2K+1}, which denotes the distance between q_{2K} and the end of the sample T, is

$$p(l_{2K+1} \mid p_{2K+1,2K+1}) = p_{2K+1,2K+1}^{l_{2K+1}}. \tag{9}$$

The marginal prior on the final regime length, l_{2K+1} is

$$p(l_{2K+1}) = \frac{\Gamma(a_{2K+1} + c_{2K+1})\Gamma(a_{2K+1} + l_{2K+1})}{\Gamma(a_{2K+1})\Gamma(a_{2K+1} + c_{2K+1} + l_{2K+1})}. \tag{10}$$

This specification results in at least some positive prior probability that $q_{2K} \geq T$ and for K breaks to occur in-sample we require $q_{2K} < T$. It is therefore important to specify the prior through the hyperparameters, a_k and c_k, such that the prior probability that $q_{2K} \geq T$ is not too large.[5] Such a specification is complicated since we entertain different values of K.

Our empirical application conducts a recursive out-of-sample forecasting exercise and thus the value of T is recursively increasing. For a given value of T our model is estimated for $g = 1, \ldots, G$ iterations in the Markov chain Monte Carlo sweep. Let $K_{T,g}$ denote the number of breaks estimated on the gth iteration when the sample size is T. We specify a_{TR}=11 and c_{TR}=2 such that the expected duration of a transition regime is $E(l_{TR}) = \frac{a_{TR}+c_{TR}-1}{c_{TR}-1}$=12. There are $K_{T,g}$ transition regimes in-sample and thus the sum of the expected prior durations of all transition regimes is $K_{T,g}E(l_{TR})$. Since there are $K_{T,g}+1$ stable regimes in-sample the prior expected duration of each is equal to $\frac{T-K_{T,g}E(l_{TR})}{K_{T,g}+1}$. We fix $c_{SR} = 2$ and compute a_{SR} on the gth iteration with a sample size of T to be consistent with the above values such that it equals $\left(\frac{T-K_{T,g}E(l_{TR})}{K_{T,g}+1}\right) - 1$. This results in an 'internally consistent' prior such that the total

[5]We are grateful to a referee for pointing this out.

prior expected duration of all regimes is equal to the sample size thereby limiting the prior probability that $q_{2K} \geq T$ (Pastor and Stambaugh, 2001).

Since our changepoint methodology differs from that of Chib (1998) insofar as it is not constructed as a Hierarchical Hidden Markov model we can marginalise the $p_{k,k}$s for simplicity.[6]

2.2.2. Prior on the number of structural breaks.

The use of Chib (1998)'s algorithm by Pastor and Stambaugh (2001) restricts their model to a geometric prior on the regime durations. Such a monotonically decreasing prior is unrealistic in many scenarios, placing more weight on numerous short regimes occurring in the sample. This could lead to a more complicated picture of the changepoint process than is supported by the data. To avoid charges of overfitting in the forecasting exercise we specify a counterbalancing geometric prior distribution over the number of breaks, K, with hyperparameter, $\lambda = 0.1$

$$p(K) = (1 - \lambda)^{K-1}\lambda. \tag{11}$$

The explanation of the choice of the value 0.1 for λ is detailed in Appendix A. Figure 1 illustrates the distribution $p(K \,|\, \sum_{k=1}^{2K+1} l_k = T)$. It demonstrates how the priors on l and K work together to give an effective prior on K in which the mass is concentrated on a number of breaks being between 10 and 14, with a spread between 1 and 32.

[6] The values of all hyperparameters used for the in-sample analysis, all of which except λ, ρ and a_{SR} are taken directly from Pastor and Stambaugh (2001), are provided in Table I. For the out-of-sample analysis hyperparameter values are recursively estimated using only the data available at the time each forecast is made.

Figure 1: Effective Prior

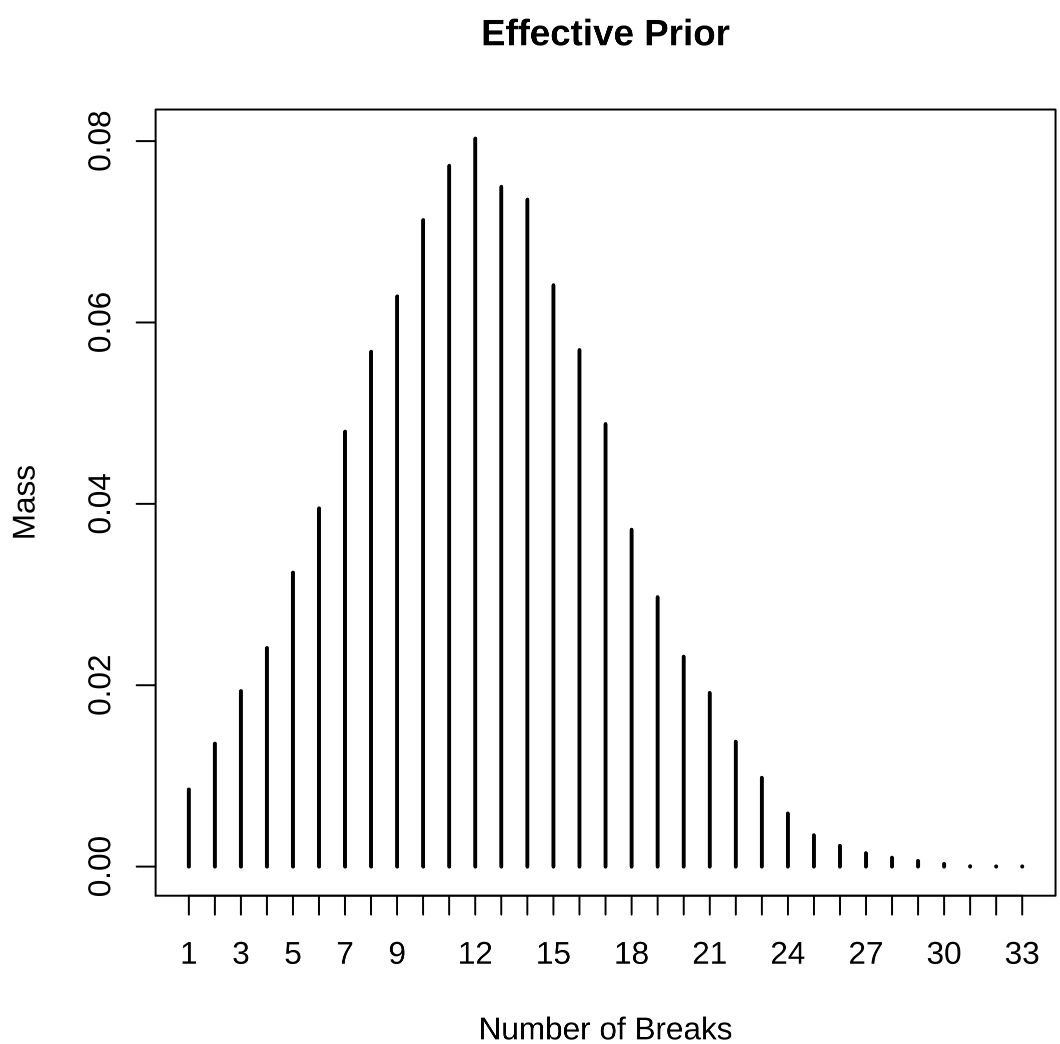

Figure 1: **Effective prior on the number of structural breaks.** This figure displays the effective prior distribution of the number of structural breaks that occur during the sample 1871 to 2013. Most of the prior mass is concentrated on the number of breaks being between 10 and 14 with a spread between 1 and 33 breaks. This specification corresponds to both the prior belief of a break occurring approximately every ten years and the findings of Pastor and Stambaugh (2001).

2.2.3. Beliefs about the TR parameters.

The prior distribution of the negative discount condition, b_j, is assumed to be normal with mean $\bar{b} < 0$ and variance σ_b^2

$$p(b_j) \propto \exp\left\{-\frac{(b_j - \bar{b})^2}{2\sigma_b^2}\right\}, \qquad j = 1, \ldots, K. \tag{12}$$

An inverse gamma distribution is placed on the prior of the volatility during TRs

$$p(\sigma_{j,j+1}^2) \propto \frac{1}{\sigma_{j,j+1}^{\eta+1}} \exp\left\{-\frac{(\eta - 2)\alpha^2}{2\sigma_{j,j+1}^2}\right\}, \qquad \sigma_{j,j+1}^2 > 0, \qquad j = 1, \ldots, K. \tag{13}$$

The values of $\bar{b}$ and σ_b^2 are recursively estimated using Campbell (1991)'s variance decomposition of aggregate returns and only the data available at the time each forecast is made.[7]

2.2.4. Beliefs about the premium's association with volatility.

The prior relation between the equity premium and volatility in a SR is specified as

$$\mu_i = \gamma\psi_i\sigma_i^2, \qquad i = 1, \ldots, K + 1, \tag{14}$$

in which $\gamma > 0$ and $\psi = (\psi_1, \ldots, \psi_{K+1})$. The prior aggregate price of risk (ratio of equity premium to variance) is represented by γ. A gamma prior distribution is placed on γ

$$p(\gamma) \propto \gamma^{a_\gamma - 1} \exp\left\{-\frac{\gamma}{b_\gamma}\right\}, \qquad \gamma > 0. \tag{15}$$

[7] We refer the reader to the Appendix of Pastor and Stambaugh (2001) for details of this specification.

The prior distribution on ψ_i is assumed to be gamma

$$p(\psi_i) \propto \psi_i^{(\nu/2)-1} \exp\left\{-\frac{\psi_i \nu}{2}\right\}, \qquad \psi_i > 0, \qquad i = 1, \ldots, K+1. \tag{16}$$

Following Pastor and Stambaugh (2001) we set $\nu = 10$ and recursively estimate α^2 based on the results in Campbell (1991) and recursively estimate a_γ and b_γ such that the prior mean of γ equals unconditional price of risk estimate (see Pastor and Stambaugh (2001) for details).

2.2.5. Beliefs about magnitudes of changes in the premium.

A hierarchical multivariate normal prior distribution with mean $\overline{\mu}$ and covariance matrix V_μ^{-1} is placed on μ, the vector of mean excess returns

$$p(\mu \mid \overline{\mu}, K) \propto \exp\left\{-\frac{1}{2}(\mu - \overline{\mu}\imath)' V_\mu^{-1} (\mu - \overline{\mu}\imath)\right\}, \qquad \mu > 0,$$

$$p(\overline{\mu}) \propto 1, \qquad \overline{\mu} > 0,$$

in which $\imath$ denotes a $(K+1) \times 1$ vector of ones. The prior on $\overline{\mu}$ is noninformative except for the positive truncation. Let Δ_i denote $\mu_{i+1} - \mu_i$. Pastor and Stambaugh (2001) state that the "unconditional prior variance of Δ_i for any i is equal to $\sigma_\Delta^2 = \mathrm{Var}(\mu_{i+1} - \mu_i)$. The value of σ_μ^2 that produces a desired value of σ_Δ is computed by simulation."

Our interpretation of their description is that all inter-regime dependencies occur through the hyperparameter $\bar{\mu}$, which is an ineffective model for producing forecasts. We therefore deviate from this and specify the equity premium values in the SRs, μ, as an AR(1) process, assuming that μ_{i+2} is conditionally independent from $\mu_{1:i}$ given μ_{i+1}, for all i. The $(K +$

1) $\times (K+1)$ covariance matrix is

$$V_\mu^{-1} = \sigma_\mu^2 \begin{pmatrix} 1 & -\rho & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1+\rho^2 & -\rho & 0 \\ 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & \cdots & 0 & -\rho & 1 \end{pmatrix}, \tag{17}$$

in which the unconditional prior variance of μ_i is denoted σ_μ^2. We set $\rho = 0.9$ and $\sigma_\mu = 0.03$.

Let the parameter vector, θ, consist of $b, \sigma_{TR}^2, \gamma, \psi_i, \mu$, and $\overline{\mu}$. All the priors except those on μ and $\overline{\mu}$ are assumed to be independent. Using (14) to substitute for the elements of σ_{SR}^2 the joint posterior distribution is

$$\begin{aligned} p(\theta \mid x, q) \propto & \left(\prod_{j=1}^{K} p(b_j)\right) \left(\prod_{j=1}^{K} p(\sigma_{j,j+1})\right) p(\gamma) \left(\prod_{i=1}^{K+1} p(\psi_i)\right) \\ & \times \left(\prod_{k=1}^{2K+1} p(l_k)\right) p(\mu \mid \overline{\mu}, K) p(\overline{\mu} \mid K) p(K) \\ & \times \left(\prod_{i=1}^{K+1} \left(\frac{\psi_i \gamma}{\mu_i}\right)^{l_{2i-1}/2}\right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\mu_i} \gamma \psi_i \right\} \\ & \times \left(\prod_{j=1}^{K} \frac{1}{\sigma_{j,j+1}^{l_{2j}}}\right) \exp\left\{ -\frac{1}{2} \sum_{j=1}^{K} \sum_{t=q_{2j-1}+1}^{q_{2j}} \frac{\left(x_t - \left(\frac{\mu_j + \mu_{j+1}}{2} + b_j \left(\mu_{j+1} - \mu_j\right)\right)^2\right)}{\sigma_{j,j+1}^2} \right\}. \end{aligned} \tag{18}$$

Table I

Hyperparameter Values

Hyperparameter	SR	TR	Regime Unspecific
a_k	$\left(\frac{T-K_{T,g}E(l_{TR})}{K_{T,g}+1}\right)-1$	11	
c_k			2
$\overline{b}$		-15.13	
σ_b		5.04	
α^2		0.00063	
η		10	
a_γ	18.7		
b_γ	0.1		
ν		10	
λ			0.1
σ_μ	0.03		
ρ	0.9		

Table 1: Hyperparameter Values. This table presents the specified values of all the hyperparameters for the various prior distributions in our multiple breaks model. Except for the values of λ and ρ, all hyperparameter values are taken from Pastor and Stambaugh (2001). The autoregressive parameter on the change in the equity premium, ρ, is specified to equal 0.9. The hyperparameter on the prior on K, λ, is set equal to 0.1 because this value corresponds to a realistic effective prior (explained in Appendix A). The value of a_k in stable regimes is dependent upon the estimated number of breaks,K_g, on the gth iteration of the MCMC loop with a sample size of T as explained in Section 2.2.1.

2.3. Posterior predictive distribution

To generate out-of-sample forecasts we simply need to simulate the model forward in time, for each set of parameters sampled in the MCMC run. The average of the predicted values then corresponds to the mean prediction when all uncertainty surrounding the parameters of the model is taken into account. Since our forecasting exercise follows Campbell and Thompson (2008) we need only consider one-step ahead forecasts and can therefore simply generate forecasts using the parameters from the final in-sample regime

$$
\begin{aligned}
\mathbb{E}(x_{T+1} \mid x_{1:T}) &= \int_K \int_\theta \mathbb{E}(x_{T+1} \mid x_{1:T}, \theta, K) p(\theta \mid x_{1:T}, K) p(K \mid x_{1:T}) d\theta dK \\
&\approx \frac{1}{N} \sum_{(n)=1}^{N} \mathbb{E}\left(x_{T+1} \mid x_{1:T}, \theta^{(n)}, K^{(n)}\right) \\
&= \frac{1}{N} \mu^{(n)}_{K^{(n)}+1},
\end{aligned}
$$

where (n) denotes the iteration of the chain.

3. Forecasting Methodology

A recursive out-of-sample forecasting exercise is conducted to compare the forecasting performance of the alternative models that allow for structural breaks considered in this paper with the benchmark null model, the historical average. Formally, the one step ahead forecast made at time T using the benchmark model is

$$x_{T+1} = \frac{1}{T} \sum_{t=1}^{T} x_t. \tag{19}$$

The first of the ad hoc alternative models is an exponential smoothing model with persistence parameter, ω. A one-step ahead forecast made at time T using the exponential smoothing model is

$$
\begin{aligned}
x_{T+1} &= \omega x_T + \omega(1-\omega) x_{T-1} + \omega(1-\omega)^2 x_{T-2} +, \ldots, +\omega(1-\omega)^{T-1} x_1 && (20) \\
&= \omega \sum_{t=0}^{T-1} (1-\omega)^t x_{T-t}. && (21)
\end{aligned}
$$

A range of values for ω are entertained in the empirical analysis and results are reported for ω=0.99 because that value delivered the strongest forecasting performance. The remaining ad hoc break models are three rolling window models with window lengths set at 5, 10, and 20 years, respectively. The one-step ahead forecast using the rolling window model is therefore

$$x_{T+1} = \frac{1}{l_w} \sum_{t=T-l_w+1}^{T} x_t, \tag{22}$$

where l_w denotes the length of the window. The out-of-sample forecasting procedure follows that of Campbell and Thompson (2008).

3.1. Data

We use monthly equity premium data, that is, the total rate of return on the aggregate stock market portfolio minus the risk-free rate. The aggregate market portfolio is the S&P 500. Returns on the S&P 500 since 1927 are sourced from CRSP, and before 1927 from Robert Shiller's website. Since 1920 the risk-free rate is the Treasury-bill rate. Prior to 1920 there was no risk-free short-term debt, and therefore we use the risk-free rate estimated by Goyal and Welch (2008). They estimate the risk-free rate from Commercial paper rates for New York City sourced from the National Bureau of Economic Research Macrohistory data base (see Goyal and Welch for a detailed description).

Since total return data before 1927 are constructed by linear interpolation of lower-frequency dividend payments and therefore may be of dubious quality we limit the out-of-sample evaluation period to the CRSP data period, but use the earlier data to estimate the models. We also need a substantial accumulation of data before we can attempt to forecast which is another reason for limiting the out-of-sample evaluation period to the CRSP data period. We present forecast results for the entire out-of-sample period as well as three subsamples corresponding to those analysed by Campbell and Thompson (2008).

3.2. Evaluation of Forecasting Performance

To evaluate the forecasting performance of the competing models we calculate the following out-of-sample R-squared statistic

$$R^2_{OS} = 1 - \frac{\sum_{t=1}^{T}(r_t - r_{A,t})^2}{\sum_{t=1}^{T}(r_t - r_{N,t})^2},$$

where $r_{N,t}$ and $r_{A,t}$ denote the forecasts from the null and alternative models, respectively, at period $t-1$. A positive R^2_{OS} value indicates the alternative model in question delivers a lower MSE than the historical average.

While our Bayesian framework is readily applicable to evaluating density forecasts, we consider only point forecasts to make our results directly comparable to the majority of the existing literature on equity premium forecasts.[8] We leave evaluation of density forecasts using a Bayesian model that admits breaks to future work.

We also examine the economic value of forecasts to a mean-variance investor. The utility calculation in the remainder of this section follows Campbell and Thompson (2008), Goyal and Welch (2008), and Rapach, Strauss, and Zhou (2009). We calculate the average utility gain for a mean-variance investor from forecasting with each of the alternative models compared with forecasting with the null model. We assume an investor has a risk-aversion coefficient, A, of three, and that the weight of the portfolio allocated to equities is restricted

[8]A handful of studies on equity premium forecasts that focus on point forecasts are Campbell and Thompson (2008), Goyal and Welch (2008), Rapach et al. (2009), and Neely, Rapach, Tu, and Zhou (2014).

to lie between 0% and 150% to prevent short-selling of equities and investors taking more than 50% leverage.

The average utility derived from a portfolio constructed using the historical average is

$$U_N = \hat{\mu}_N - \frac{1}{2} A \hat{\sigma}_N^2, \tag{23}$$

where $\hat{\mu}_N$ and $\hat{\sigma}_N^2$ denote the sample mean and variance for the return on the portfolio formed using the null model over the out-of-sample period, and A denotes the risk aversion coefficient. To calculate the portfolio return from the null model we must calculate the proportion of the portfolio to be allocated to stocks at period t

$$w_{N,t} = \frac{1}{A} \frac{r_{N,\mathrm{t+1}}}{\hat{\sigma}_{t+1}^2}, \tag{24}$$

where $r_{N,t+1}$ denotes the equity premium forecast from the null model for $t+1$ made using information available at time t and $\hat{\sigma}_{t+1}^2$ denotes the variance of stock returns estimated using a 5-year rolling window of lagged monthly stock returns, as used by Campbell and Thompson (2008).

The average utility for the same investor using the forecasts from the alternative model is

$$U_A = \hat{\mu}_A - \frac{1}{2} A \hat{\sigma}_A^2, \tag{25}$$

where $\hat{\mu}_A$ and $\hat{\sigma}_A^2$ denote the sample mean and variance for the return on the portfolio formed using the alternative model over the out-of-sample period. The proportion of the portfolio that shall be allocated to stocks at period t when forecasting with the alternative model is

$$w_{A,t} = \frac{1}{A} \frac{r_{A,\mathrm{t+1}}}{\hat{\sigma}^2_{t+1}}, \tag{26}$$

where $r_{A,\mathrm{t+1}}$ denotes the equity premium forecast using the alternative model for $t+1$ made using information available at time t.

The utility gain is measured as $U_A - U_N$ and this figure is multiplied by 1200 for monthly returns and 100 for annual returns to represent it as an average annualised percentage return. The utility gain can be viewed as the portfolio management fee an investor would be willing to pay to exploit the information available in the alternative model compared with having only the information from the null model.

Any economic gains delivered by a forecasting model may dissipate if the asset allocation exercise has high turnover and thus incurs high transaction costs. We therefore evaluate the utility gain net of transaction costs. Following Neely et al. (2014) (see also Balduzzi and Lynch (1999)) monthly turnover is the percentage of wealth traded every month and transaction costs are computed using monthly turnover while assuming proportional transactions cost equal to 50 basis points per transaction. We evaluate the utility gain net of transaction costs.

4. Empirical results

First, we revisit the full sample estimation of the equity premium conducted by Pastor and Stambaugh (2001) to explore whether fixing the number of breaks at 15 leads to any economically meaningful estimation error of the equity premium relative to allowing the number of breaks to vary and subsequently incorporating the uncertainty surrounding the number of breaks into the estimates of the premium. Second, we explore whether allowing for breaks in a mean model can improved forecasting performance.

4.1. Full sample estimates of the equity premium

We now revisit the work of Pastor and Stambaugh (2001) exploring whether fixing the number of breaks equal to 15 relative to incorporating the uncertainty surrounding the number of breaks into the estimates of the equity premium leads to any considerable estimation error. Figure 2 graphs the evolution of the equity premium estimated from both methods. Of course, both estimates follow similar paths but the exact differences between the two are worthy of note. First, fixing the number of breaks tends to underestimate the premium for the majority of the sample. Second, as can be seen in Figure 3 the estimation error is often as high as one percentage point and around the time of the Wall Street Crash, which exhibits the greatest degree of uncertainty in our sample, the estimation error is above 2.5 percentage points. The estimation error is also notably large around the oil price shocks of the 1970s and the Global Financial Crisis. The root-squared difference in estimates averages 0.78% across the sample as displayed in Figure 4.

Estimation errors of 0.78% and at times as much as 2.5% could have considerable economic significance. The equity premium is one of the most important numbers in finance being used for estimating the cost of capital, the CAPM, and allocating wealth between risky and risk-free portfolios. Our procedure could therefore have far reaching economic significance for many applications in finance.

4.1.1. Mixing of the algorithm

To compare the mixing of our algorithm with those of Chib (1998) and Koop and Potter (2007) we report the popular convergence diagnostic of Gelman and Rubin (1992) for all three algorithms. Since the equity premium estimate μ is the parameter of interest both in-sample and in the out-of-sample analysis that follows, we focus on this parameter when computing the Gelman-Rubin statistic. To calculate this statistic, for each method we estimate $m = 1, \ldots, M$ independent Markov chains for $g = 1, \ldots G$ iterations. We first compute the variance of each chain[9]

$$s_m^2 = \frac{1}{G} \sum_{g=1}^{G} (\mu_{gm} - \bar{\mu}_m)^2 \tag{27}$$

and second the variance pooled across chains

$$W = \frac{1}{M} \sum_{m=1}^{M} s_m^2. \tag{28}$$

Letting

$$\bar{\bar{\mu}} = \frac{1}{M} \sum_{m=1}^{M} \bar{\mu}_m \tag{29}$$

and

$$B = \frac{G}{M-1} \sum_{m=1}^{M} (\bar{\mu}_m - \bar{\bar{\mu}})^2 \tag{30}$$

the Gelman-Rubin diagnostic is computed as

$$R = \sqrt{\frac{\hat{Var}(\mu)}{W}} \tag{31}$$

[9] Calculations are averaged across regimes, the subscript notation for which is omitted here for expositional simplicity.

in which

$$\hat{Var}(\mu) = (1 - 1/G)W + \frac{1}{G}B. \tag{32}$$

We carry out the same procedure for the methods of Chib (1998) and Koop and Potter (2007).

The closer the Gelman-Rubin statistic R is to one the more evidence there is that the algorithm has converged. The statistic from our algorithm is equal to 1.02 while Chib (1998) delivers a value of 1.04 and Koop and Potter (2007) 1.06. All the values are relatively close to one suggesting good mixing of each algorithm. This evidence suggests that our algorithm is most likely to have converged with the algorithm of Koop and Potter (2007) least likely. This may reflect the sensitivity of the Hidden Markov Models to poor chain initialisation, whereas the the reversible jump Markov chain Monte Carlo algorithm can more easily escape from this (Fearnhead, 2006).

4.1.2. Computational issues

Estimating Bayesian models with multiple changepoints can incur large computational costs. We compare run times for our algorithm with the algorithms of Chib (1998) and Koop and Potter (2007). Table II displays the results. The computational burden of our method relative to the Hidden Markov Models is reduced. The method of Koop and Potter (2007), denoted KP, is clearly the most burdensome with run times that are approximately twice as long as our reversible jump (RJ) algorithm. The algorithm of Chib (1998) improves upon Koop and Potter (2007) but still takes longer than our method to complete.

Table II
Computation times

Method	Number of iterations					
	50,000	100,000	250,000	500,000	1m	3m
RJ	1	2	6	14	30	96
Chib	2	3	10	20	42	132
KP	2	5	13	26	57	181

Table II: Computation times. This table displays computation times in minutes of the model presented here estimated using our algorithm and the algorithms of Chib (1998) and Koop and Potter (2007) for various numbers of iterations. Run times are rounded to the nearest minute.

4.2. Out-of-sample forecasts of the equity premium

We now explore whether allowing for structural breaks can improve upon the out-of-sample performance of a mean model when forecasting the equity premium at both the monthly and annual horizons. We recursively estimate the hyperparameter values when making each forecast. Given our initial estimation period is so long (1871-1926), the recursively estimated hyperparameters are relatively stable thereafter as displayed in Figures 5-7.

Figure 2: Full sample estimates of the equity premium

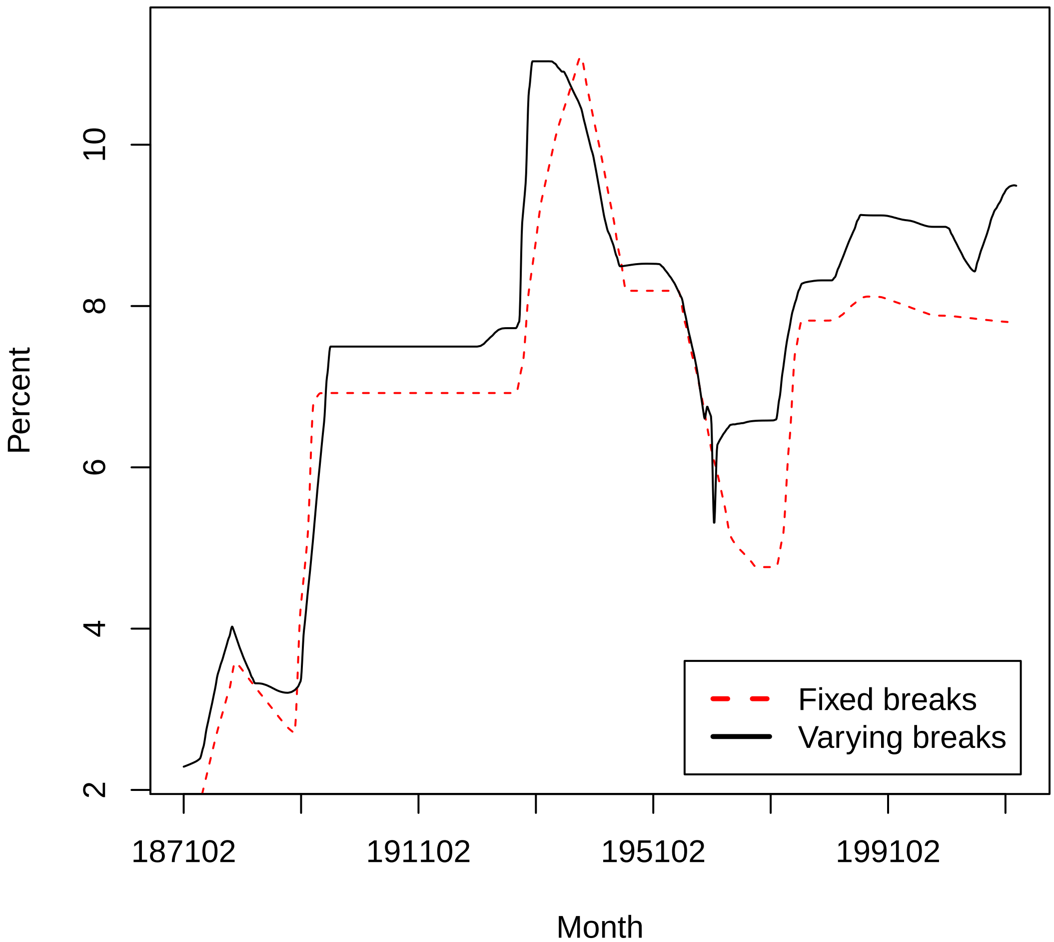

Figure 2: Full sample estimates of the equity premium. This figure graphs the evolution of the equity premium over the full sample estimated from the model which allows the number of breaks to vary (black solid line) and the model which fixes the number of breaks at 15 (red dotted line).

Figure 3: Difference in equity premium estimates

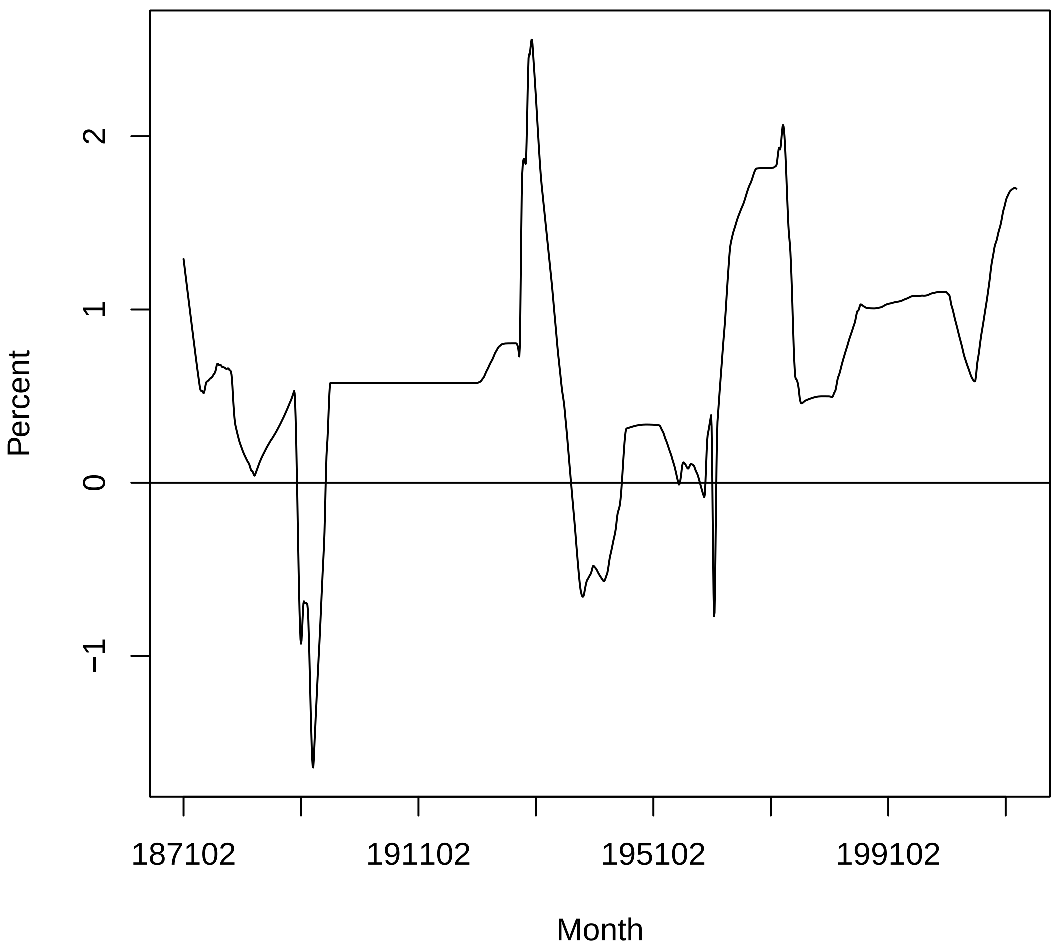

Figure 3: Difference in equity premium estimates. This figure graphs the difference over the full sample in the equity premium estimates from the model which allows the number of breaks to vary and the model which fixes the number of breaks at 15.

Figure 4: Root squared difference in equity premium estimates

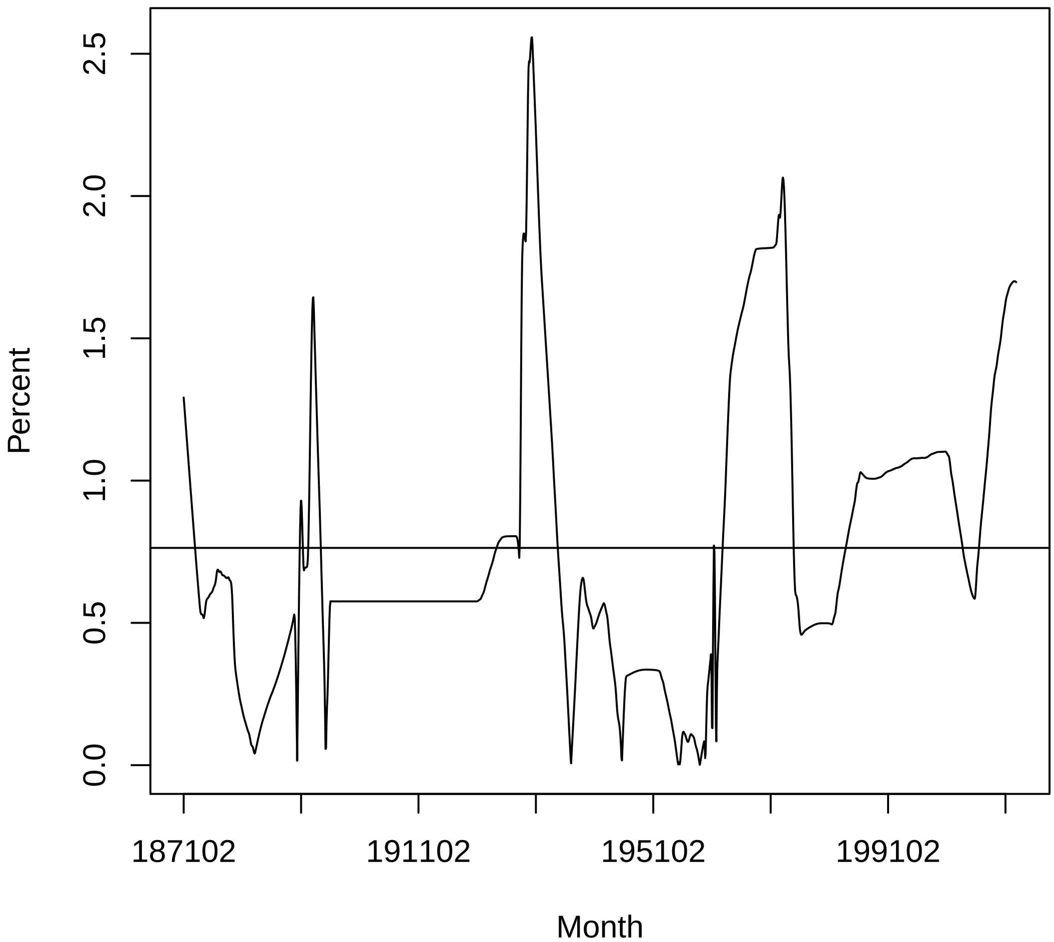

Figure 4: Root squared difference in equity premium estimates. This figure graphs the root squared difference over the full sample in the equity premium estimates from the model which allows the number of breaks to vary and the model which fixes the number of breaks at 15.

Figure 5: Hyperparameter estimates

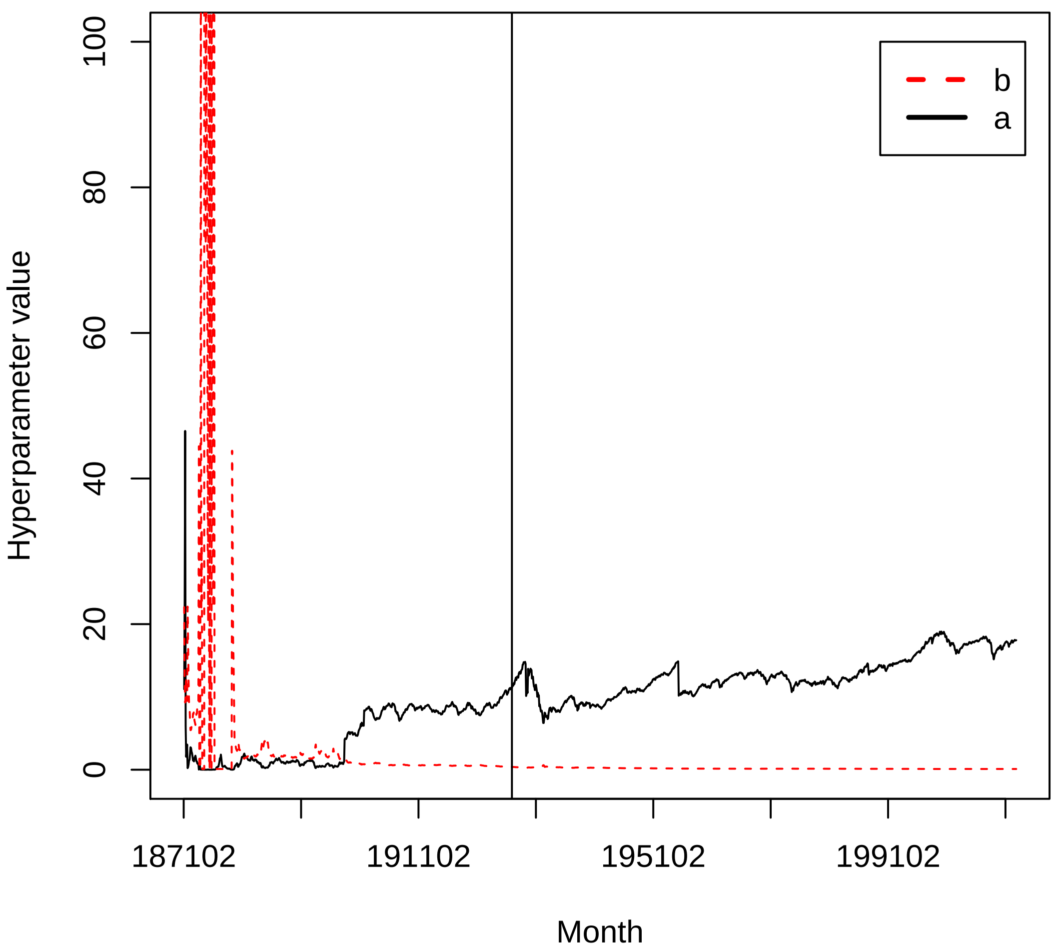

Figure 5: Hyperparameter estimates. This figure graphs the recursively estimated hyperparameter values a_γ (black solid line) and b_γ (red dotted line). The vertical line marks the date at which the out-of-sample forecasting period begins, January 1927.

Figure 6: Hyperparameter estimates

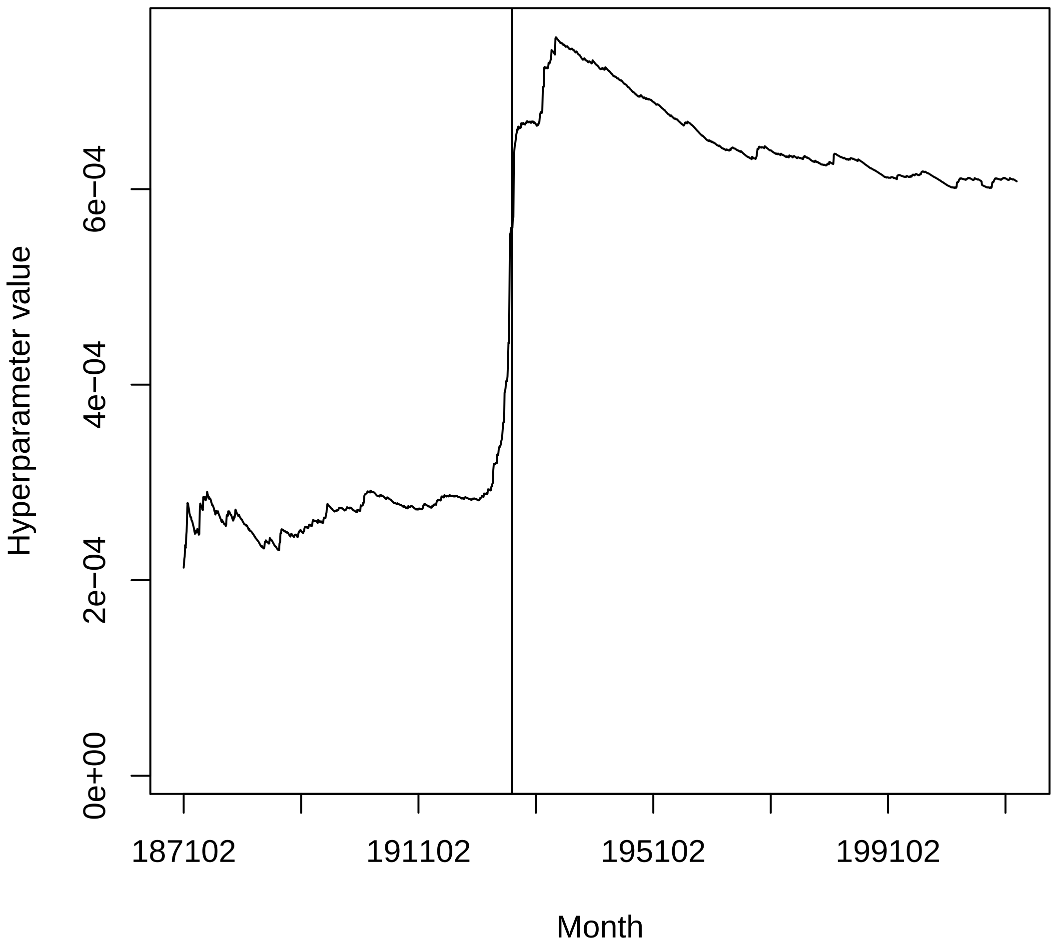

Figure 6: Hyperparameter estimates. This figure graphs the recursively estimated hyperparameter value α^2 over the sample. The vertical line marks the date at which the out-of-sample forecasting period begins, January 1927.

Figure 7: Hyperparameter estimates

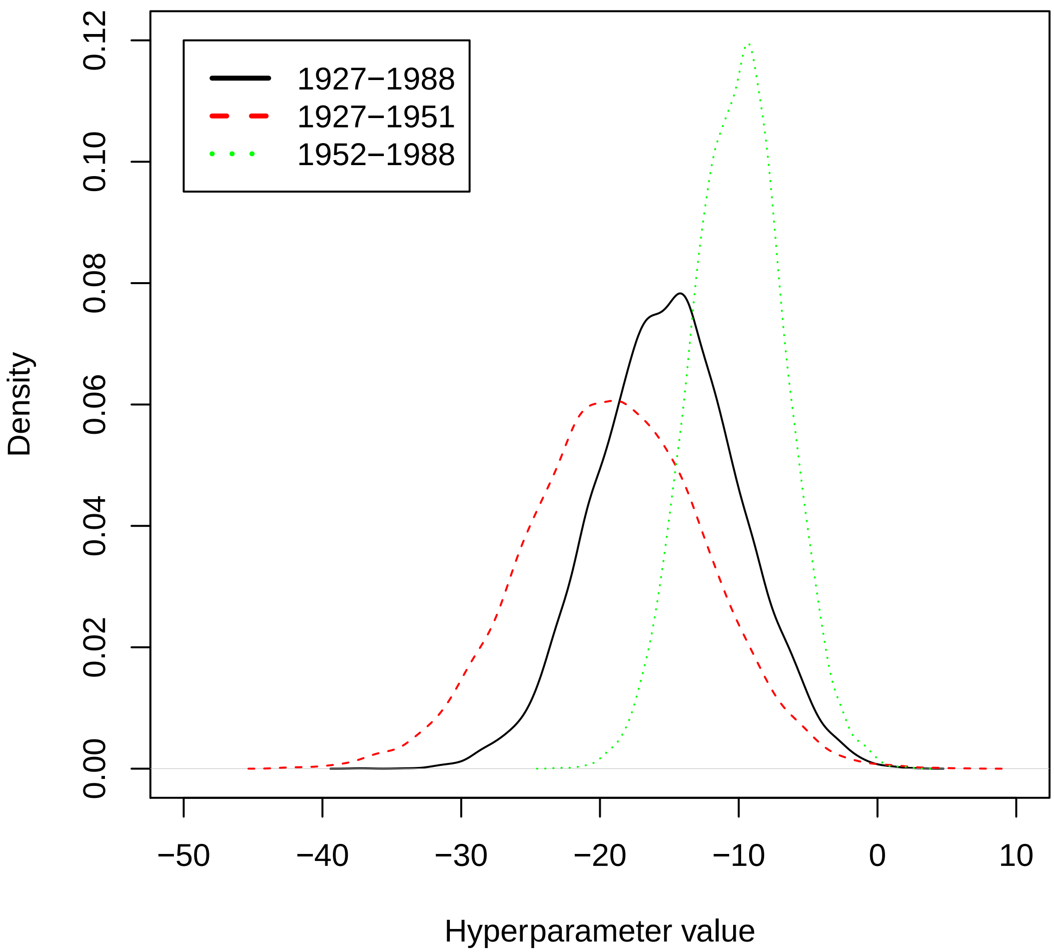

Figure 7: Hyperparameter estimates. This figure plots the prior distribution of b_j using three sets of hyperparameters $\bar{b}$ and σ_b. The black plot uses the values computed from Campbell (1991)'s variance decomposition of stock returns from 1927 through 1988. The red dashed and green dotted lines are computed using the same technique using data from 1927 through 1951, and 1951 through 1988, respectively.

4.2.1. Monthly prediction

Tables III and IV display the performance of the alternative models that allow for breaks in forecasting the equity premium compared with the null model, the historical average. Specifically, they display both the out-of-sample R-squared and the utility gain values. It can be seen in Table III that the R^2_{OS} statistics for all of the simple univariate models that allow for breaks are negative over the entire out-of-sample period, indicating inferior performance relative to the historical average. The historical average also outperforms the simple univariate break models over the three subsamples apart from a few exceptions. This poor performance is not surprising in the light of the reservations we expressed above about these methods.

It is important, as Goyal and Welch (2008) note, that any full sample outperformance is not clustered around one or two short periods. It is also important that any overall success is just as evident in recent forecasts. We see that the multiple breaks model satisfies these criteria. It delivers consistent outperformance with a positive R-squared value and a positive annualised utility gain in each of the three subsamples.

Table III

Forecasts of Monthly Equity Premium

	Sample									
Model	1927-2013				1927-1956		1956-1980		1980-2013	
	R^2_{OS}	ΔU	$Turn$	Net ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU
5-year Rolling Window	-2.24	-1.01	3.74	-2.47	-2.64	-2.37	0.19	2.15	-2.73	-2.07
10-year Rolling Window	-0.49	-0.10	2.86	-0.63	-0.34	-1.24	0.17	2.41	-1.29	-0.89
20-year Rolling Window	-0.29	-0.28	2.02	-0.39	0.13	0.17	-1.03	0.57	-0.94	-1.36
Exponential Smoothing	-0.44	-0.13	2.33	-0.56	-0.32	-1.11	-0.39	1.26	-0.83	-1.22
Multiple Breaks	0.75	1.37	2.95	1.22	0.57	0.74	1.57	2.86	0.65	1.02
Multiple Breaks (pre)	0.34	0.79	1.55	0.69	0.26	0.61	0.82	1.57	0.32	0.74

Table III: Forecasts of monthly equity premium. This table presents out-of-sample error statistics for forecasts of the monthly equity premium. Data prior to January 1927 are from Robert Shiller's website. A positive value of R^2_{OS} indicates the model has superior out-of-sample forecasting power than the null model. The null model is the historical average, except for Multiple Breaks (pre) in which the alternative model is our model and the null model is the model with a prespecified number of breaks. R^2_{OS} values have been multiplied by 100. ΔU denotes the difference in average utility derived from forecasting with the alternative model compared with the null model. ΔU is reported as an annualised percentage return. The utility function is $E(R_p) - \frac{A}{2}Var(R_p)$, the risk aversion coefficient, A, is set equal to 3, and R_p denotes the return on the portfolio. $Turn$ denotes the average turnover of the portfolio exercise from the alternative model divided by that of the null model. Net ΔU denotes the difference in average utility after accounting for transaction costs.

We graph in Figure 8 the cumulative R-squared over the out-of-sample period based on forecasts from each of the univariate models that allow for breaks compared with forecasts from the historical average. An increase in the line indicates the alternative model outperformed the null over the period in question and a fall in the line indicates the null outperformed the alternative model. Figure 8 shows a general upward drift in the cumulative R-squared based on the multiple breaks model relative to the historical average from just

Figure 8: Cumulative R-Squared Based on Forecasts of Monthly Equity Premium

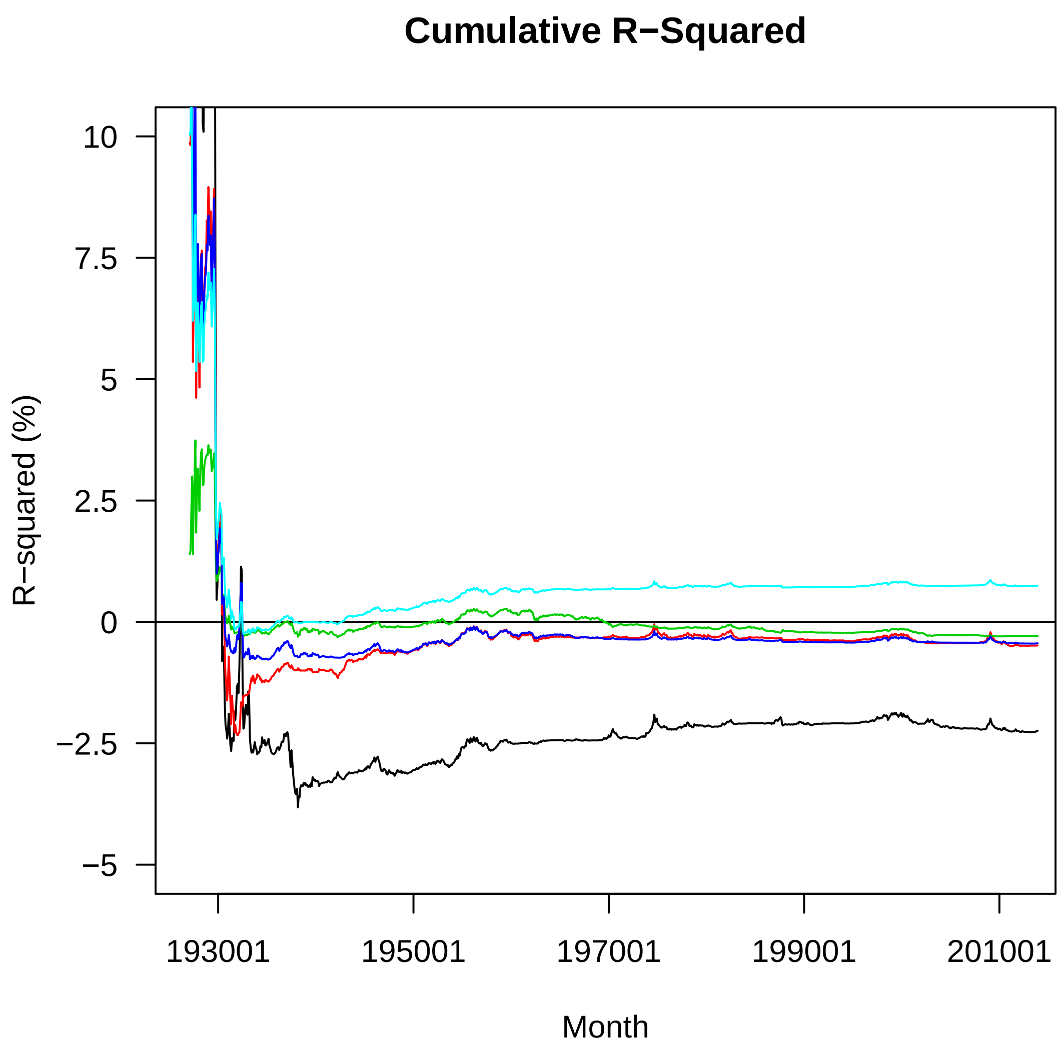

Figure 8: Cumulative R-squared based on forecasts of the monthly equity premium from univariate break models. Following Campbell and Thompson (2008), this figure displays the cumulative R-squared over the out-of-sample period for each of the univariate models considered in this paper that account for breaks compared with the historical average when forecasting the monthly equity premium.

after 1930 through 2013 that is not clustered around one or two short periods.

The multiple breaks model developed in this paper outperforms the historical average over the entire out-of-sample period, 1927 to 2013, delivering an R-squared value of 0.75% and an annualised utility gain of 1.37%. The value of ΔU given in Table III determines the annual percentage portfolio management fee an investor would have been willing to pay to have exploited the information in each of the alternative models compared with having only the information provided by the historical average. Campbell and Thompson (2008) suggest that even a small value for ΔU of 0.5% could lead to the investor profiting from the additional information provided by the alternative model. We see that a mean-variance investor could have profited from forecasting with the multiple breaks model compared with the historical average.

We further compare the performance of our forecasts with a method that prespecifies the number of breaks and is estimated by the algorithm of Chib (1998). Forecasting in the presence of breaks requires careful management of bias-variance tradeoffs (Pesaran and Timmermann, 2007). Determining the number of breaks endogenously may induce large estimation error into the parameter estimates and therefore lead to poor forecasts. The bottom panel of Table III - Multiple Breaks (pre) - displays the out-of-sample R-squared and utility gains delivered by our model relative to the model that prespecifies the number of breaks. Our method outperforms this model in an economically meaningful manner with an R-squared value of 0.34% and an annualised utility gain of 0.79%. This economically meaningful outperformance holds across all subsamples and also holds after accounting for transaction costs. This suggests that while allowing for a prespecified number of breaks is important, obtaining the full benefits from breaks requires determining the number of breaks endogenously.

The full sample result that the investor with the specified utility function and trading rule would have paid a 1.4% annual premium to have benefited from using the multiple breaks model is economically meaningful when measured against an average value for the equity premium of 6.3%.

If the multiple breaks model induces sufficiently high turnover in the portfolio allocation exercise transaction costs may eliminate the utility gain relative to the historical average. Table III reports that the average monthly turnover of the portfolio obtained from the multiple breaks model divided by the average turnover from the historical average is 2.95. The utility gain net of transaction costs is 1.22% and while lower than the original value of 1.37% is still economically meaningful and thus transaction costs have not eliminated the utility gain derived from the improved forecasts.

4.2.2. Annual prediction

We now turn to evaluating forecasts of the annual equity premium. The idea of occasional stochastic breaks does not imply any particular natural length of a single period measured in calendar time. We obtain annual forecasts from the model estimated using annual returns, which are measured as compounded monthly returns. We follow the procedure of Campbell and Thompson (2008) and evaluate overlapping monthly forecasts of the annual premium. Each month we construct forecasts of the equity premium over the next twelve months. So for the model that we use for forecasting each month we estimate the model using 12 months of data compounded up to that same month each year. So for example for forecasts made on March 1st all the monthly data used to estimate the model is compounded over 12 months ending on February 28th each year and so on.

Working with overlapping data allows us to fully exploit the data to estimate the economic value of these 12 month forecasts. This advantage comes at a price, since assessing the statistical significance of the goodness of fit requires bootstrapping to accommodate the overlapping data (see e.g. Mark, 1995 and Kilian, 1999). Bootstrapping is impossibly costly in the recursive fitting exercise deployed here, but the alternative would be to work with non-overlapping annual returns, resulting in a reduction in the number of forecasts by a factor of 12, and laying open the possibility of data snooping by carefully selecting a month in which to make the annual predictions. Therefore, although we report the R-squared values of our predictions, we make no claims on their statistical significance. Instead we focus on whether any difference in performance is economically meaningful.

We see in Table IV that the multiple breaks model outperforms the historical average for the entire out-of-sample period, delivering an R-squared value of 7.94% and this outperformance is economically meaningful with a utility gain of 1.11%. It is also more successful than all the alternative approaches that allow for structural breaks. As with monthly forecasts we see consistent outperformance of the multiple breaks model compared with the historical mean across all three subsamples.

Our method also outperforms the model with a prespecified number of breaks when forecasting the annual equity premium delivering an R-squared value of 3.92% and an annualised utility gain of 0.82%. The outperformance is economically meaningful across all subsamples. In line with the results for monthly forecasts, allowing for a prespecified number of breaks improves forecasts but determining the number of breaks endogenously is required to derive the full benefits of allowing for breaks in the forecasting exercise.

The average turnover of the portfolio constructed from the multiple breaks model divided by the average turnover of the portfolio from the historical average is equal to 2.72 ($Turn$). The higher turnover from the multiple breaks model relative to the historical average reduces the utility gain from 1.11% to 1.02% (Net ΔU) and thus the improved forecasts are still economically meaningful after accounting for transaction costs.

Table IV
Forecasts of Annual Equity Premium

	Sample									
Model	1927-2013				1927-1956		1956-1980		1980-2013	
	R^2_{OS}	ΔU	$Turn$	Net ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU
5-year Rolling Window	-8.27	-0.29	3.88	-0.52	-11.87	-1.66	17.71	3.27	-14.89	-1.79
10-year Rolling Window	2.69	0.34	3.13	0.20	4.82	-0.51	7.73	1.93	-7.31	-0.06
20-year Rolling Window	-0.15	-0.09	2.15	-0.14	4.34	0.53	-9.53	0.52	-7.54	-1.25
Exponential Smoothing	3.34	0.40	2.97	0.32	4.64	0.18	3.96	1.38	-1.10	-0.19
Multiple Breaks	7.94	1.11	2.72	1.02	7.85	1.06	13.35	2.03	4.53	0.71
Multiple Breaks (pre)	3.92	0.82	1.32	0.78	3.33	0.80	5.29	0.98	3.04	0.60

Table IV: Forecasts of annual equity premium. This table presents out-of-sample error statistics when forecasting the simple annual equity premium. Forecasts are based on 12 overlapping annual returns per year. Data prior to January 1927 are from Robert Shiller's website. A positive value of R^2_{OS} indicates the model has superior out-of-sample forecasting power than the null model. The null model is the historical average, except for Multiple Breaks (pre) in which the alternative model is our model and the null model is the model with a prespecified number of breaks. R^2_{OS} values have been multiplied by 100. ΔU denotes the difference in average utility derived from forecasting with the alternative model compared with the null model. ΔU is reported as an annualised percentage return. The utility function is $E(R_p) - \frac{A}{2}Var(R_p)$, the risk aversion coefficient, A, is set equal to 3, and R_p denotes the return on the portfolio. $Turn$ denotes the average turnover of the portfolio obtained from the alternative model divided by the average turnover of the portfolio from the null model. Net ΔU denotes the difference in average utility net of transaction costs.

Inspection of the subsample results in Table IV highlights the importance of the consistency criterion. The 5-year rolling window model substantially outperforms all other models in the 1956 to 1980 subsample, including the multiple breaks model, but performs badly in both of the other subsamples. Both the exponential smoothing model and the 10-year rolling

Figure 9: Cumulative R-Squared Based on Forecasts of Annual Equity Premium

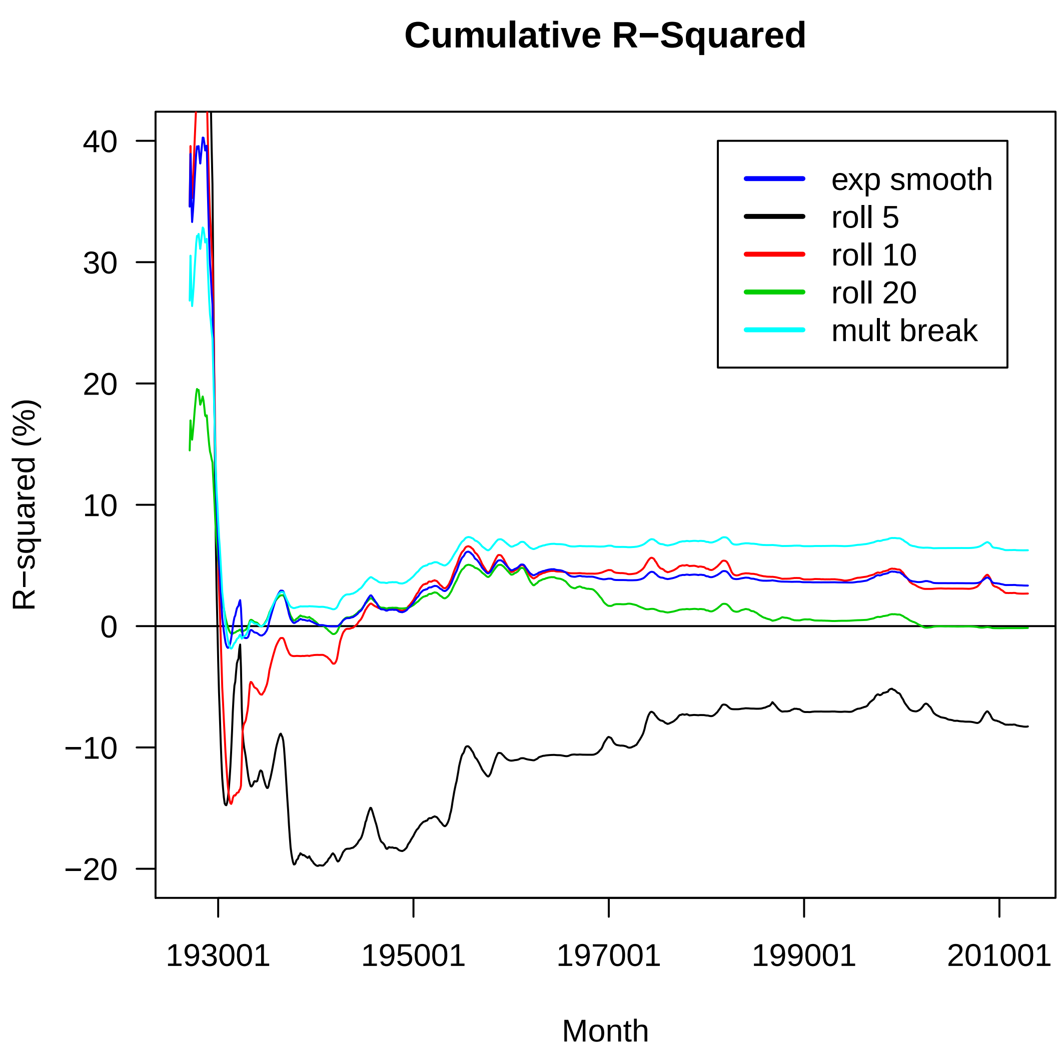

Figure 9: Cumulative R-squared based on forecasts of the annual equity premium from univariate break models. Following Campbell and Thompson (2008), this figure displays the cumulative R-squared over the out-of-sample period for each of the univariate models considered in this model that account for breaks compared with the historical average when forecasting the simple annual equity premium.

window beat the historical mean over the full sample measured both by R-squared and the utility gain. However, the success of both is largely due to the 1956 to 1980 subsample, and crucially both underperform in the most recent sample. These results highlight how long a period we need in order to judge the success of a proposed model for forecasting the equity premium. In some contexts 25 years might seem a reasonable horizon over which to judge the success of a model, but any investor who started in 1980 to trade on the basis of a model that would have been profitable over the previous 25 years would have lost money over the next 33 years, and a risk averse investor with the preferences and trading rule specified here would have lost utility.

In Figure 9 we graph the year by year forecasting performance of the different models at the annual horizon. Figure 9 shows that the success of the 5-year rolling window in the middle subsample is due to two short periods in the 1950s and the 1970s. On the other hand the multiple breaks model performs strongly from the early 1930s through the end of the sample.

Dangl and Halling (2012) and Johannes, Korteweg, and Polson (2014) report that allowing for time-variation in a predictive regression using popular variables such as the dividend yield leads to improved equity premium forecasts relative to the historical average. We find, however, that allowing for permanent shifts in a mean model outperforms the historical average when forecasting the equity premium at both the monthly and annual horizons. It would therefore be interesting to investigate whether allowing for time-variation in a predictive regression using the popular predictors outperforms a model which allows for time-variation in the mean. If it does not outperform the time-varying mean model then these predictive variables may not exhibit any additional predictive power. We leave a comprehensive investigation of this matter to future work.

Figure 10: Posterior Distribution of the Number of Breaks

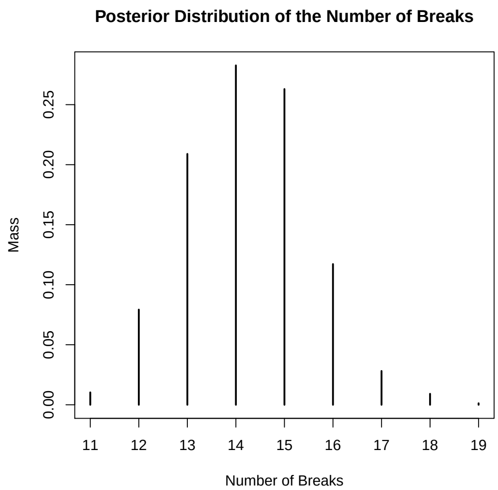

Figure 10: Posterior distribution of the number of breaks. This plot displays the posterior distribution of the number of breaks that occur over the entire sample (1871 to 2013) when predicting the monthly equity premium. The mode number of breaks is 14 with approximately 80% of the mass lying between 12 and 16 breaks.

4.3. Distribution of structural breaks

Figure 10 displays the posterior distribution of the number of structural breaks, K, that occur over the sample 1871 to 2013 when estimating the monthly equity premium. The posterior mode is 14 breaks with a probability of 27%. A posterior mode of 14 breaks corresponds to a break occurring approximately every 10 years on average which corroborates the findings of Pastor and Stambaugh (2001). Recall the effective prior (see Figure 1) has a mode number of breaks of 12 with a spread between 1 and 29 breaks. The posterior distribution of K has a spread from 11 to 19 breaks, with approximately 80% of the mass lying between 12 and 16 breaks.

Each structural break is assumed to occur gradually and is thus represented by a TR which separates two SRs. Figure 11 displays the posterior probability that a TR begins in any given month over the out-of-sample period, 1927 to 2013. The greatest posterior break

Figure 11: Posterior Distribution of Changepoint Locations

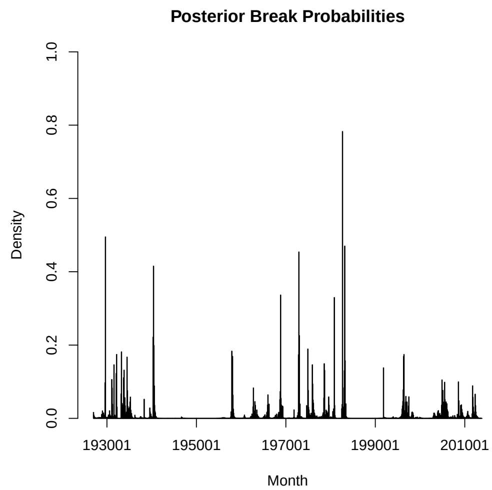

Figure 11: Posterior distribution of the changepoint locations. This plot displays the posterior distribution of the changepoint locations that occur in the out-of-sample period, 1927-2013, when predicting the monthly equity premium. The most likely changepoint location, with a posterior break probability of 84%, is November 1982.

probability of 84% is attributed to November 1982. There are 25 months that have a posterior break probability of at least 20%. There is a 76% probability of a TR beginning in either October or November 1929. There is a 93% probability of a TR beginning between June and September 1940, and an 84% probability of a TR beginning between December 1972 and February 1973. More recently, there is a 65% probability of a TR beginning between December 1991 and February 1992. Finally, the recent financial crisis appears to represent a structural break, with an 82% probability of a TR beginning between October 2007 and April 2008.

4.4. Sensitivity of results to priors

Finally, we explore the sensitivity of the out-of-sample forecasting results to the hyperparameter value choices that control the exact shapes of the prior distributions. Specifically, we analyse the ability of our multiple breaks method relative to the historical average to forecast the monthly equity premium when we adjust one hyperparameter value at a time. We report out-of-sample R-squared values and the economic utility gain in terms of annualised percentage for the full sample and across subsamples. Table V displays the results.

For all but one case, the out-of-sample R-squared values remain positive over the full sample and all subsamples while the utility gain is economically meaningful. The only case in which the outperformance breaks down is when we set $\sigma_\mu = 100$. The higher is σ_μ the more prior weight is placed on large shifts in the equity premium. Pastor and Stambaugh (2001) specified a relatively tight prior here because large shifts in the equity premium are implausible on economic grounds. Furthermore, forecasting in the presence of structural breaks requires careful management of bias-variance tradeoffs (Pesaran and Timmermann, 2007). Increasing σ_μ to 100 results in (i) implausible equity premium estimates and (ii) very noisy estimates causing all improvement in forecasting power relative to the benchmark model to dissipate. This result is in line with the literature. For example, Wachter and Warusawitharana (2009) report that allowing the prior variance on the dividend yield slope coefficient in a predictive regression of stock returns to become very large eliminates all improved forecasting power.

Table V

Sensitivity of results to priors

	Sample							
Hyperparameter	1927-2013		1927-1956		1956-1980		1980-2013	
	R^2_{OS}	ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU	R^2_{OS}	ΔU
$a_\gamma = 1$	0.69	1.23	0.61	0.80	1.47	2.80	0.75	1.13
$\alpha^2 = 0.1$	0.80	1.47	0.63	0.75	1.68	2.88	0.59	0.98
$\eta = 1$	0.76	1.33	0.59	0.74	1.60	2.93	0.62	1.00
$\bar{b}$ = -5	0.69	1.28	0.54	0.68	1.56	2.80	0.60	1.01
$\rho = 0.5$	0.80	1.42	0.60	0.75	1.69	2.91	0.59	0.97
$\nu = 5$	0.74	1.35	0.55	0.69	1.57	2.85	0.69	1.04
$a_{SR} = 450$	0.78	1.40	0.58	0.74	1.61	2.90	0.66	1.02
$\sigma_\mu = 0.01$	0.71	1.33	0.54	0.69	1.53	2.69	0.72	1.11
$\sigma_\mu = 0.1$	0.77	1.38	0.60	0.77	1.56	2.82	0.65	1.06
$\sigma_\mu = 100$	-1.29	-1.89	-0.87	-1.48	-1.87	-3.06	-1.32	-1.98

Table V: Sensitivity of results to priors. This table presents out-of-sample error statistics when forecasting the simple monthly equity premium. We adjust one hyperparameter at a time and report the out-of-sample R-squared and improvement in economic utility for the full sample and across subsamples for a mean-variance investor from forecasting with the multiple break model relative to using only the information available in the prevailing mean.

5. Conclusions

In this paper we have developed a Bayesian multiple changepoint model that determines the number of structural breaks endogenously in a generic and flexible framework that is computationally efficient. Our approach is applicable to many highly structured models for which some existing approaches are computationally more cumbersome. One particular highly structured and inherently attractive model is that introduced by Pastor and Stambaugh (2001) that estimates the equity premium over a long sample in the presence of a fixed number of breaks. We estimate this model using our multiple changepoint approach that determines the number of breaks endogenously. The computational efficiency of our approach enables the model to be easily applied in a recursive forecasting exercise of the U.S. equity premium. Goyal and Welch (2008) report that the historical average outperforms a range of popular economic predictors when forecasting the equity premium. We find that the forecasting performance of the univariate approach can be further improved by allowing for an unknown number of breaks in a theory-informed model of the equity premium. For mean-variance investors, our approach delivers forecasts that are an economically meaningful improvement compared to use of the historical average.

Appendix A. Specifying the Prior on the Number of Structural Breaks

The RJMCMC algorithm is not restricted to a geometric prior on the regime durations. Any prior can be used in a straightforward manner without complicating the computation. With a sufficiently diffuse prior our method can determine any number of breaks ranging from zero, assuming a constant equity premium, to a break every period as in the exponential smoothing model. However, we have kept the geometric prior here to remain as close as possible to the model of Pastor and Stambaugh (2001). Note, however, the effective prior is more complicated because it is also affected by the geometric prior placed on the number of breaks in the model (see Section 2.2.2.2), which in itself is also not a particularly restrictive prior belief. To investigate this issue further, the sampler was run without data (so the likelihood contribution is always set to 1). To find the value of λ that corresponded to the prior belief of a break occurring approximately every ten years, we sampled from the prior distribution of K using many different values for λ between 0 and 1.

Appendix B. Estimating the Model

Due to the complex nature of our Bayesian model, the joint posterior distribution is not analytically tractable. We use simulation techniques known as Markov chain Monte Carlo (MCMC) methods to sample from the joint posterior. With enough simulations our samples give a good approximation to the joint posterior. The sampling of our model consists of three stages: sampling the parameters conditional on the number of changepoints and their locations, providing the local adjustment to the changepoint locations conditional on the number of changepoints, and finally sampling the number of changepoints.

Updating the parameters conditional on the changepoints: The Metropolis Hastings and Gibbs steps are used to obtain draws of the parameter vector θ conditional on K and q. Parameters are repeatedly drawn from their full conditional distributions. Draws beyond a transient burn-in stage are thinned at a specified interval to combat autocorrelation of the draws.

Following the exposition of Pastor and Stambaugh (2001) we perform the change of variables $\lambda \equiv 1/\gamma$ and $\phi_k^2 \equiv 1/\psi_k$ and define $\Delta_i \equiv \mu_{i+1} - \mu_i$. Incorporating the change of variables, the full conditional distributions are derived from the joint posterior distribution:

$$\overline{\mu} \mid \cdot \sim \mathcal{N}\left(\frac{\imath' V_\mu^{-1}\mu}{\imath V_\mu^{-1}\imath}, \frac{1}{\imath' V_\mu^{-1}\imath}\right), \qquad \overline{\mu} > 0, \tag{B1}$$

$$\phi_i^2 \mid \cdot \sim \frac{\nu + \sum\limits_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t-\mu_i)^2}{\lambda\mu_i}}{\chi^2_{\nu+l_{2i-1}}}, \qquad i = 1, \ldots, K+1, \tag{B2}$$

$$\lambda \mid \cdot \sim \frac{\frac{2}{b_\gamma} + \sum\limits_{i=1}^{K+1} \sum\limits_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t-\mu_i)^2}{\lambda\mu_i}}{\chi^2_{2a_\gamma + \sum_{i=1}^{K+1} l_{2i-1}}}, \tag{B3}$$

$$\sigma^2_{j,j+1} \mid \cdot \sim \frac{(\eta-2)\alpha^2 + \sum\limits_{t=q_{2j-1}+1}^{q_{2j}} \left(x_t - \left(\frac{\mu_j+\mu_{j+1}}{2} + b_j\Delta_j\right)\right)^2}{\chi^2_{\eta+l_{2j}}}, \qquad j = 1, \ldots, K \tag{B4}$$

$$b_j \mid \cdot \sim \mathcal{N}(m_{b_j}\nu_{b_j}, \nu_{b_j}), \qquad j = 1, \ldots, K, \tag{B5}$$

where

$$m_{b_j} = \frac{\bar{b}}{\sigma_b^2} + \frac{\Delta_j l_{2j}}{\sigma_{j,j+1}^2}\left(\frac{\sum\limits_{t=q_{2j-1}+1}^{q_{2j}} x_t}{l_{2j}} - \frac{\mu_j + \mu_{j+1}}{2}\right),$$

$$\nu_{b_j} = \frac{1}{\frac{1}{\sigma_b^2} + \frac{\Delta_j^2 l_{2j}}{\sigma_{j,j+1}^2}}.$$

The full conditional distribution of μ is

$$\begin{aligned} p(\mu \mid \cdot) \propto & \left(\prod_{i=1}^{K+1} \mu_i^{(-l_{2i-1/2})}\right) \\ & \times \exp\left\{-\frac{1}{2}\left[(\mu - \overline{\mu}\imath)' V_\mu^{-1}(\mu - \overline{\mu}\imath) + \sum_{i=1}^{K+1} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} \frac{(x_t - \mu_i)^2}{\lambda \phi_i^2 \mu_i}\right.\right. \\ & \left.\left. + \sum_{j=1}^{K} \sum_{t=q_{2j-1}+1}^{q_{2j}} \frac{\left(x_t - \left(\frac{\mu_j+\mu_{j+1}}{2} + b_j\Delta_j\right)\right)^2}{\sigma_{j,j+1}^2}\right]\right\}, \qquad \mu > 0, \\ & \propto \left(\prod_{i=1}^{K+1} \mu_i^{(-l_{2i-1/2})}\right) \exp\left\{-\frac{1}{2}\left[\mu'(V_\mu^{-1} + D_1 + D_2)\mu\right.\right. \\ & \left.\left. - 2\mu'(V_\mu^{-1}\overline{\mu}\imath + w + g) + d'\overline{x}^2\right]\right\}, \qquad \mu > 0, \end{aligned} \tag{B6}$$

where d is a $(K+1)$-vector whose ith element is

$$d_i = \frac{l_{2i-1}}{\lambda \mu_i \phi_i^2}, \tag{B7}$$

w is a $(K+1)$-vector whose ith element is

$$w_i = -\frac{l_{2i-1}}{2\lambda \phi_i^2}, \tag{B8}$$

$\overline{x}^2$ is a $(K+1)$-vector whose ith element is

$$\overline{x}_i^2 = \frac{1}{l_{2i-1}} \sum_{t=q_{2i-2}+1}^{q_{2i-1}} x_t^2, \tag{B9}$$

$$D_1 = \begin{pmatrix} z_{1,1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & z_{2,1}+z_{1,2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & z_{2,2}+z_{1,3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & z_{2,K-1}+z_{1,K} & 0 \\ 0 & 0 & 0 & \cdots & 0 & z_{2,K} \end{pmatrix} \tag{B10}$$

$$D_2 = \begin{pmatrix} 0 & -z_{3,1} & 0 & \cdots & 0 & 0 \\ -z_{3,1} & 0 & -z_{3,2} & \cdots & 0 & 0 \\ 0 & -z_{3,2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -z_{3,K} \\ 0 & 0 & 0 & \cdots & -z_{3,K} & 0 \end{pmatrix} \tag{B11}$$

$$g = \begin{pmatrix} z_{4,1} \\ z_{5,1} + z_{4,2} \\ z_{5,2} + z_{4,3} \\ \vdots \\ z_{5,K-1} + z_{4,K} \\ z_{5,K} \end{pmatrix} \tag{B12}$$

and, for $j = 1, \ldots, K$,

$$z_{1,j} = \frac{l_{2j}\left(b_j - \frac{1}{2}\right)^2}{\sigma^2_{j,j+1}} \tag{B13}$$

$$z_{2,j} = \frac{l_{2j}\left(b_j + \frac{1}{2}\right)^2}{\sigma^2_{j,j+1}} \tag{B14}$$

$$z_{3,j} = \frac{l_{2j}\left(b_j - \frac{1}{2}\right)\left(b_j + \frac{1}{2}\right)}{\sigma^2_{j,j+1}} \tag{B15}$$

$$z_{4,j} = \frac{l_{2j}\left(b_j - \frac{1}{2}\right)}{\sigma^2_{j,j+1}}\overline{x}_{2j} \tag{B16}$$

$$z_{5,j} = \frac{l_{2j}\left(b_j + \frac{1}{2}\right)}{\sigma^2_{j,j+1}}\overline{x}_{2j} \tag{B17}$$

$$\overline{x}_{2j} = \frac{1}{l_{2j}} \sum_{t=q_{2j-1}+1}^{q_{2j}} x_t. \tag{B18}$$

The Gibbs Sampler is used to draw from all of the above full conditional distributions except for that of μ, which is not a standard distribution due to the truncation at 0. We update μ one component at a time using a random walk Metropolis-Hastings step, using univariate normal proposals with variance 0.0004, which has been tuned to achieve reasonable acceptance rates.

Updating the changepoint locations: Pastor and Stambaugh (2001) recognise that using local adjustments of the changepoint locations will not allow sufficient exploration of the posterior distribution. They therefore make use of Chib (1998)'s algorithm to sample a complete new set of changepoint locations on each MCMC iteration. However, we will introduce a mechanism for introducing and removing changepoints, rendering local adjustments sufficient here and thereby avoiding the large computational burden associated with Chib (1998)'s algorithm when the time series becomes long. Even improved versions of Chib (1998)'s algorithm such as Pesaran, Pettenuzzo, and Timmermann (2006), Koop and Potter (2007), and Giordani and Kohn (2008) are prohibitively expensive because they address the full time series on every iteration. In contrast, local updates only address each changepoint and since the number of changepoints is significantly less than the length of the time series a computational saving is achieved.

Specifically, the method we use is a simple random-walk Metropolis Hastings step. For each changepoint, k, we propose to perturb its location, q_k, by a number sampled uniformly from the interval $[-5, 5]$. Each proposal is accepted with probability equal to $\min(1, \alpha)$, in which $\alpha = p(\hat{q} \mid x, \theta)/p(q \mid x, \theta)$ and $\hat{q}$ denotes the perturbed changepoint locations.

Updating the number of changepoints: The reversible jump Markov chain Monte Carlo (RJMCMC) algorithm we develop in this article relaxes the restriction that K must be fixed in advance by the user. The algorithm jumps between different numbers of changepoints, drawing the posterior $p(K \mid x)$ in the process. Determining the number of breaks endogenously and providing automated online Bayesian Model Averaging in such a way removes any requirement for offline calculations. Online Bayesian Model Averaging is important for a recursive out-of-sample forecasting exercise, in which the number of changepoints for each estimation period must be inferred automatically from the data.

RJMCMC allows an MCMC sampler to jump between statistical models of varying size (Green, 1995). In the current example, it allows the sampler to jump between models with different numbers of changepoints. On each iteration, the algorithm attempts to either add (birth move) or remove (death move) one structural break. The proposed move is then accepted with a probability which is calculated to ensure that detailed balance across the full posterior (including K) is maintained. On each iteration, a birth move is attempted with probability $b_k = 0.5$ and a death move with probability $d_k = 0.5$. Clearly, $d_0 = 0$ and $b_{k_{max}} = 0$, where k_{max} corresponds to a break occurring every period.

Birth move: If a birth move is chosen, an attempt will be made to increase K to $K+1$ and split an existing SR_i into three regimes: two new SRs separated by a TR. Let these three regimes be denoted SR^*_-, TR^*, and SR^*_+, respectively. Let the existing SR_i, which we attempt to split, be denoted SR_{i^c}. We sample SR_{i^c} uniformly from $SR_{1:K+1}$. The length of TR^*, denoted l_{TR^*}, is then proposed from a Negative Binomial distribution

$$l_{TR^*} \sim NB(p, r), \tag{B19}$$

where $p = 0.7$ and $r = 0.05$.[10] The first of the two new proposed changepoints, denoted q_{TR^*-1}, is drawn uniformly on the interval $(q_{2i^c-1}+1, q_{2i^c-1}-l_{TR^*})$, where $q_{2i^c-1}+1$ and q_{2i^c-1} denote the first and final time points of SR_{i^c}, respectively. The move is immediately rejected if $l^*_{TR} \geq l_{2i^c-1}-1$. The second proposed changepoint, denoted by q_{TR^*}, is computed as $q_{TR^*-1}+l_{TR^*}$. We now have three regimes where we once had a single SR so we must propose a new equity premium value for each of the two new SRs, new price of risk values to calculate the variance in the new SRs, the variance in the TR, and the negative discount condition to calculate the equity premium in the newly proposed TR.[11]

Let $\mu_{SR^*_-}$ and $\mu_{SR^*_+}$ denote the values of the equity premium in SR^*_- and SR^*_+, respectively. Their values are proposed from a normal distribution with mean, $\overline{x}$, equal to the mean of the equity premium data from the entire sample. The variance of this proposal distribution, $\sigma^2_{\mu^*}$, is selected to ensure the distribution has slightly heavier tails than the empirical distribution of the equity premium data from the entire sample

$$\mu_{SR^*_-}, \mu_{SR^*_+} \sim \mathcal{N}(\overline{x}, \sigma^2_{\mu^*}), \tag{B20}$$

where $\overline{x} = \sum_{t=1}^{T} x_t/T$ and $\sigma^2_{\mu^*} = 0.01$. We use an independent proposal distribution rather than a split move to increase the acceptance rate of the proposed moves. The default split move of Green (1995) often proposes equity premium values that are implausible and subsequently rejected by the data. Our proposal distribution is centred at the mean of the data and we find adopting this proposal distribution increases the acceptance probability.

The proposal distribution for the negative discount condition for the newly proposed TR,

[10]This proposal distribution was chosen to closely reflect the empirical distribution of the TR lengths observed in an initial run with no RJMCMC step incorporated.

[11]Table VI lists the variables that need to be proposed.

b_{TR^*}, is a normal distribution with hyperparameters $\overline{b^*}$ and $\sigma^2_{b^*}$

$$b_{TR^*} \sim \mathcal{N}(\overline{b}^*, \sigma^2_{b^*}), \tag{B21}$$

where

$$\overline{b^*} = \frac{\overline{x}_{TR^*} - \frac{\mu_{SR^*_+} - \mu_{SR^*_-}}{2}}{\mu_{SR^*_+} - \mu_{SR^*_-}}, \tag{B22}$$

$\overline{x}_{TR^*}$ is the mean of the excess return data from the TR^* sample, and $\sigma^2_{b^*}$ is set equal to 25.[12] Recall from (5) that the mean excess return during a TR is specified as

$$\frac{\mu_j + \mu_{j+1}}{2} + b_j(\mu_{j+1} - \mu_j), \qquad j = 1, \ldots, K. \tag{B23}$$

We simply equate expression (B23) to $\overline{x}_{TR^*}$ to create the mean of the proposal distribution, $\overline{b}^*$. We then calculate the equity premium for TR^* as

$$\mu_{TR^*} = \frac{\mu_{SR^*_-} + \mu_{SR^*_+}}{2} + b_{TR^*}(\mu_{SR^*_+} - \mu_{SR^*_-}). \tag{B24}$$

Now $\phi^2_{SR^*_-}$ and $\phi^2_{SR^*_+}$ can be proposed from their full conditional distribution. The variances for the proposed SRs are calculated as described in (14) and the variance for TR^*, $\sigma^2_{TR^*,TR^*+1}$, is proposed from its full conditional distribution.[13]

The set of new parameters required for the birth move is now complete. Although we have a complex model, we need only independently propose three parameters to add the extra regimes in the birth move. To calculate the acceptance probability we need to also know the corresponding reverse proposal density. We therefore delay calculation of the acceptance

[12] This value is chosen to be slightly larger than the variance of the empirical distribution of the output of b values from an initial run that incorporated no RJMCMC step.

[13] In which $\psi_i \equiv 1/\phi^2_i$.

probability until after describing the death move.

Death move: The death move attempts to reduce K to K-1 and thereby replace three regimes, denoted SR^c_-, TR^c, and SR^c_+, respectively, with a single SR, denoted SR^*. First, TR^c is sampled uniformly from $TR_{1:K}$. The length of SR^* is computed as

$$l_{SR^*} = l_{SR^c_-} + l_{TR^c} + l_{SR^c_+}. \tag{B25}$$

The values of b_{TR^c} and $\sigma^2_{TR^c,TR^c+1}$ are discarded, as are $\mu_{SR^c_-}$, $\mu_{SR^c_+}$, $\phi^2_{SR^c_-}$, and $\phi^2_{SR^c_+}$. The μ_{SR^*} value is proposed using (B20) and the new $\phi^2_{SR^*}$ is proposed from its full conditional distribution. The full set of parameters that require proposing for both the birth and death moves are presented in Table VI.

Table VI

Proposed Parameters

SR	TR
Birth move	
$\mu_{SR^*_-}$	
$\mu_{SR^*_+}$	
$\phi^2_{SR^*_-}$	
$\phi^2_{SR^*_+}$	
	b_{TR^*}
	$\sigma^2_{TR^*,TR^*+1}$
	l_{TR^*}
Death move	
μ_{SR^c}	
$\phi^2_{SR^c}$	

Table VI: Proposed parameters. This table presents the parameters that require proposing in the birth and death moves, respectively.

Acceptance probability: The model-jumping move is simply a Metropolis Hastings move, and therefore an acceptance probability needs to be calculated to decide whether or not to accept the proposed new values. Denote by θ_K the set of within-model parameters for model K; the distribution from which we wish to sample is proportional to

$$p(K)p(\theta_K \mid K)p(x \mid \theta_K, K).$$

Green (1995), and Green and Hastie (2012) demonstrate that to retain reversibility of the chain the moves from (K, θ_k) should be considered as the following process:

- Propose a model K^* into which to jump, with probability $j(K, K^*)$
- Propose an appropriate random vector u with density q_{K,K^*}
- Calculate $(\theta^*_{K^*}, u^*) = h_{K,K^*}(\theta_K, u)$ where h_{K,K^*} is a diffeomorphism
- Accept the proposal with probability

$$\max\left\{1, \frac{p(K^*)p(\theta^*_{K^*} \mid K^*)p(x \mid \theta^*_{K^*}, K^*)}{p(K)p(\theta_K \mid K)p(x \mid \theta_K, K)} \frac{j(K^*, K)q_{K^*,K}(u^*)}{j(K, K^*)q_{K,K^*}(u)} \left|\frac{\partial(\theta^*, u^*)}{\partial(\theta, u)}\right|\right\}, \tag{B26}$$

otherwise reject.

This description of the methodology has proved to be a barrier for many interested in applying the method. However some generic simplifications make the method much simpler.

The key simplification we propose is to assume a limited kind of dependency between θ_K and $\theta^*_{K^*}$. As in the proposals described above, much of θ_K is retained unchanged in $\theta^*_{K^*}$ (i.e. all the parts of the state vector not corresponding to the regime that is being split or merged). The new terms in $\theta^*_{K^*}$ should be drawn either from independent proposal distributions (such as the fixed Gaussian distribution used for $\mu^*_\pm$) or from simple distributions conditional only on the parts of θ_K that are to be retained, and other newly proposed parts of $\theta^*_{K^*}$ (such as the full conditionals used to propose $\phi^2_{SR^*_\pm}$).

Considering u and u^* to be vectors of independent uniform random variables, and the diffeomorphism h_{K,K^*} to consist of identity functions and inverse cumulative distribution functions, it is a simple calculation in this instance to show that the acceptance probability in (B26) becomes

$$\max\left\{1, \frac{p(K^*)p(\theta^*_{K^*} \,|\, K^*)p(x \,|\, \theta^*_{K^*}, K^*)}{p(K)p(\theta_K \,|\, K)p(x \,|\, \theta_K, K)} \frac{j(K^*, K)}{j(K, K^*)} \frac{q(\theta^*_{K^*}, \theta_K)}{q(\theta_K, \theta^*_{K^*})}\right\} \tag{B27}$$

where $q(\theta_K, \theta^*_{K^*})$ is the sampling density of all the new components of $\theta^*_{K^*}$ when jumping from θ_K and, similarly, $q(\theta^*_{K^*}, \theta_K)$ is the sampling density of all the new components of θ_K when jumping from $\theta^*_{K^*}$.

In our example, whenever a move is proposed we have $j(K, K^*) = j(K^*, K) = \frac{1}{2}$ (except at the boundaries of model space, where we may have $j = 0$ or $j = 1$). Furthermore, since a jump always either introduces or removes a single TR, we need only consider two proposal densities. The proposal density when the birth of a new TR is proposed is

$$\begin{aligned} q(\theta_K, \theta^*_{K+1}) =& \frac{1}{K+1} p_{NB(p,r)}(l^*_{TR^*}) f_{\bar{x},\sigma^2_{\mu^*}}(\mu^*_{SR^*_-}) f_{\bar{x},\sigma^2_{\mu^*}}(\mu^*_{SR^*_+}) f_{\overline{b_{TR}},\sigma^2_{b_{TR}}}(b^*_{TR^*}) \\ & \times p(\sigma^{2*}_{TR^*,TR^*+1} \,|\, \cdots) p(\phi^{2*}_{SR^*_-} \,|\, \cdots) p(\phi^{2*}_{SR^*_+} \,|\, \cdots) \end{aligned}$$

where the $(K+1)^{-1}$ corresponds to the uniform selection of one of the current stable regimes to break, $p_{NB(p,r)}$ is the mass function for the negative binomial proposal distribution for the length of the new transition regime, f_{m,s^2} gives the density of a normal random variable with mean m and variance s^2, and $p(X \,|\, \cdots)$ is the full conditional distribution of the random variable X. The proposal density for a death proposal is

$$q(\theta_{K+1}, \theta_K^*) = \frac{1}{K+1} f_{\bar{x}, \sigma^2_{\mu^*}}(\mu_{SR^*}) p(\phi^2_{SR^*} \mid \cdots)$$

where in this instance the $(K+1)^{-1}$ corresponds to the uniform selection of one of the $K+1$ transition regimes to kill.

The use of the full conditional distributions results in canceling in (B27), so that values of parameters proposed through these proposals have no effect on the acceptance probability. The model-jumping step therefore has fewer effective dimensions and the chance of accepting a proposal is increased.

The chain is run for 3 million iterations with a burn-in of 250,000 and a thinning interval of 20. Pastor and Stambaugh (2001) run 606,000 iterations and since we are also estimating the number of breaks we have a larger parameter space to traverse. Running 3 million iterations is computationally feasible in our framework since each iteration is much faster than the forward-backward approach of Chib (1998) chosen by Pastor and Stambaugh (2001). This number of iterations allows the sampler to traverse the entire parameter space. For each set of priors and hyperparameters we run 2 chains with different seeds and starting values to diagnose convergence.

REFERENCES

Andrews, Donald WK, 1993, Tests for parameter instability and structural change with unknown change point, *Econometrica: Journal of the Econometric Society* 821–856.

Ang, Andrew, and Geert Bekaert, 2002, Regime switches in interest rates, *Journal of Business & Economic Statistics* 20, 163–182.

Bai, Jushan, 1997, Estimating multiple breaks one at a time, *Econometric Theory* 13, 315–352.

Bai, Jushan, and Pierre Perron, 1998, Estimating and testing linear models with multiple structural changes, *Econometrica* 47–78.

Bai, Jushan, and Pierre Perron, 2003, Computation and analysis of multiple structural change models, *Journal of applied econometrics* 18, 1–22.

Balduzzi, Pierluigi, and Anthony W Lynch, 1999, Transaction costs and predictability: Some utility cost calculations, *Journal of Financial Economics* 52, 47–78.

Bauwens, Luc, Gary Koop, Dimitris Korobilis, and Jeroen VK Rombouts, 2015, The contribution of structural break models to forecasting macroeconomic series, *Journal of Applied Econometrics* 30, 596–620.

Bec, Frédérique, Anders Rahbek, and Neil Shephard, 2008, The acr model: a multivariate dynamic mixture autoregression, *Oxford Bulletin of Economics and Statistics* 70, 583–618.

Campbell, John Y, 1991, A variance decomposition for stock returns, *The Economic Journal* 101, 157–179.

Campbell, John Y, and Samuel B Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average?, *Review of Financial Studies* 21, 1509–1531.

Carlin, Bradley P, Alan E Gelfand, and Adrian FM Smith, 1992, Hierarchical bayesian analysis of changepoint problems, *Applied statistics* 389–405.

Chen, Bin, and Yongmiao Hong, 2012, Testing for smooth structural changes in time series models via nonparametric regression, *Econometrica* 80, 1157–1183.

Chernoff, Herman, and Shelemyahu Zacks, 1964, Estimating the current mean of a normal distribution which is subjected to changes in time, *The Annals of Mathematical Statistics* 999–1018.

Chib, Siddhartha, 1998, Estimation and comparison of multiple change-point models, *Journal of Econometrics* 86, 221–241.

Clements, Michael, and David Hendry, 1998, *Forecasting economic time series* (Cambridge University Press).

Dangl, Thomas, and Michael Halling, 2012, Predictive regressions with time-varying coefficients, *Journal of Financial Economics* 106, 157–181.

Fearnhead, Paul, 2006, Exact and efficient bayesian inference for multiple changepoint problems, *Statistics and computing* 16, 203–213.

Fox, Emily B, Erik B Sudderth, Michael I Jordan, and Alan S Willsky, 2011, A sticky hdp-hmm with application to speaker diarization, *The Annals of Applied Statistics* 1020–1056.

Gelman, Andrew, and Donald B Rubin, 1992, Inference from iterative simulation using multiple sequences, *Statistical science* 457–472.

Giacomini, Raffaella, and Barbara Rossi, 2009, Detecting and predicting forecast breakdowns, *The Review of Economic Studies* 76, 669–705.

Giordani, Paolo, and Robert Kohn, 2008, Efficient bayesian inference for multiple changepoint and mixture innovation models, *Journal of Business & Economic Statistics* 26.

Giordani, Paolo, Robert Kohn, and Dick van Dijk, 2007, A unified approach to nonlinearity, structural change, and outliers, *Journal of Econometrics* 137, 112–133.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Green, Peter J, 1995, Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika* 82, 711–732.

Hamilton, James D, 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the Econometric Society* 357–384.

Inoue, Atsushi, and Barbara Rossi, 2011, Identifying the sources of instabilities in macroeconomic fluctuations, *Review of Economics and statistics* 93, 1186–1204.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *The Journal of Finance* 69, 611–644.

Kilian, Lutz, 1999, Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses, *Review of Economics and Statistics* 81, 652–660.

Kim, Chang-Jin, James C Morley, and Charles R Nelson, 2005, The structural break in the equity premium, *Journal of Business & Economic Statistics* 23.

Ko, Stanley IM, Terence TL Chong, Pulak Ghosh, et al., 2015, Dirichlet process hidden markov multiple change-point model, *Bayesian Analysis* 10, 275–296.

Koop, Gary, and Simon Potter, 2010, A flexible approach to parametric inference in nonlinear and time varying time series models, *Journal of Econometrics* 159, 134–150.

Koop, Gary, and Simon M Potter, 2007, Estimation and forecasting in models with multiple breaks, *The Review of Economic Studies* 74, 763–789.

Lundbergh, Stefan, Timo Teräsvirta, and Dick Van Dijk, 2003, Time-varying smooth transition autoregressive models, *Journal of Business & Economic Statistics* 21, 104–121.

Maheu, John M, and Thomas H McCurdy, 2009, How useful are historical data for forecasting the long-run equity return distribution?, *Journal of Business & Economic Statistics* 27, 95–112.

Mark, Nelson C, 1995, Exchange rates and fundamentals: Evidence on long-horizon predictability, *The American Economic Review* 201–218.

Neely, Christopher J, David E Rapach, Jun Tu, and Guofu Zhou, 2014, Forecasting the equity risk premium: the role of technical indicators, *Management Science* 60, 1772–1791.

Pastor, Lubos, and Robert F. Stambaugh, 2001, The equity premium and structural breaks, *The Journal of Finance* 56, 1207–1239.

Pesaran, M Hashem, Davide Pettenuzzo, and Allan Timmermann, 2006, Forecasting time series subject to multiple structural breaks, *The Review of Economic Studies* 73, 1057–1084.

Pesaran, M Hashem, and Allan Timmermann, 2002, Market timing and return prediction under model instability, *Journal of Empirical Finance* 9, 495–510.

Pesaran, M Hashem, and Allan Timmermann, 2007, Selection of estimation window in the presence of breaks, *Journal of Econometrics* 137, 134–161.

Rapach, David E, Jack K Strauss, and Guofu Zhou, 2009, Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* 821–862.

Song, Yong, 2014, Modelling regime switching and structural breaks with an infinite hidden markov model, *Journal of Applied Econometrics* 29, 825–842.

Stephens, DA, 1994, Bayesian retrospective multiple-changepoint identification, *Applied Statistics* 159–178.

Stock, James H, and Mark W Watson, 1996, Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics* 14, 11–30.

Wachter, Jessica A, and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.